Graduate Texts in Mathematics

Harold Davenport

Multiplicative Number Theory

Second Edition

Springer-Verlag Berlin Heidelberg GmbH

Graduate Texts in Mathematics 74

Harold Davenport

Multiplicative Number Theory

Second Edition

Revised by Hugh L. Montgomery

Springer-Verlag
Berlin Heidelberg GmbH

Harold Davenport
(Deceased)
Cambridge University
Cambridge
England

Hugh L. Montgomery
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109
USA

AMS Subject Classification (1980): 10-01, 10Hxx

Library of Congress Cataloging in Publication Data

Davenport, Harold, 1907–1969
Multiplicative number theory.

(Graduate texts in mathematics; 74)
Revised by Hugh Montgomery.
Bibliography: p.
Includes index.
1. Numbers, Theory of. 2. Numbers, Prime.
I. Montgomery, Hugh L. II. Title. III. Series.
QA241.D32 1980 512'.7 80-26329

The first edition of this book was published by Markham Publishing Company, Chicago, IL, 1967.

Originally published by Springer-Verlag Berlin Heidelberg New York in 1980.
Softcover reprint of the hardcover 2nd edition 1980

9 8 7 6 5 4 3 2 1

ISBN 978-1-4757-5929-7 ISBN 978-1-4757-5927-3 (eBook)
DOI 10.1007/978-1-4757-5927-3

CONTENTS

PREFACE TO THE SECOND EDITION

Although it was in print for a short time only, the original edition of *Multiplicative Number Theory* had a major impact on research and on young mathematicians. By giving a connected account of the large sieve and Bombieri's theorem, Professor Davenport made accessible an important body of new discoveries. With this stimulation, such great progress was made that our current understanding of these topics extends well beyond what was known in 1966. As the main results can now be proved much more easily. I made the radical decision to rewrite §§23–29 completely for the second edition. In making these alterations I have tried to preserve the tone and spirit of the original.

Rather than derive Bombieri's theorem from a zero density estimate for L functions, as Davenport did, I have chosen to present Vaughan's elementary proof of Bombieri's theorem. This approach depends on Vaughan's simplified version of Vinogradov's method for estimating sums over prime numbers (see §24). Vinogradov devised his method in order to estimate the sum $\sum_{p \leqslant X} e(p\alpha)$; to maintain the historical perspective I have inserted (in §§25, 26) a discussion of this exponential sum and its application to sums of primes, before turning to the large sieve and Bombieri's theorem.

Before Professor Davenport's untimely death in 1969, several mathematicians had suggested small improvements which might be made in *Multiplicative Number Theory*, should it ever be reprinted. Most of these have been incorporated here; in particular, the nice refinements in §§12 and 14, were suggested by Professor E. Wirsing. Professor L. Schoenfeld detected the only significant error in the book, in the proof of Theorems 4 and 4A of §23. Indeed these theorems are false as they stood, although their corollaries, which were used later, are true. In considering the extent and nature of my revisions, I have benefited from the advice of Professors Baker, Bombieri, Cassels, Halberstam, Hooley, Mack, Schmidt, and Vaughan, although the responsibility for the decisions taken is entirely my own. The assistance throughout of Mrs. H. Davenport and Dr. J. H. Davenport has been invaluable. Finally, the

mathematical community is indebted to Professor J.-P. Serre for urging Springer-Verlag to publish a new edition of this important book.

H.L.M.

PREFACE TO THE FIRST EDITION

My principal object in these lectures was to give a connected account of analytic number theory in so far as it relates to problems of a multiplicative character, with particular attention to the distribution of primes in arithmetic progressions. Most of the work is by now classical, and I have followed to a considerable extent the historical order of discovery. I have included some material which, though familiar to experts, cannot easily be found in the existing expositions.

My secondary object was to prove, in the course of this account, all the results quoted from the literature in the recent paper of Bombieri[1] on the average distribution of primes in arithmetic progressions; and to end by giving an exposition of this work, which seems likely to play an important part in future researches. The choice of what was included in the main body of the lectures, and what was omitted, has been greatly influenced by this consideration. A short section has, however, been added, giving some references to other work.

In revising the lectures for publication I have aimed at producing a readable account of the subject, even at the cost of occasionally omitting some details. I hope that it will be found useful as an introduction to other books and monographs on analytic number theory.

§§23 and 29 contain recent joint work of Professor Halberstam and myself, and I am indebted to Professor Halberstam for permission to include this. The former gives our version of the basic principle of the large sieve method, and the latter is an average result on primes in arithmetic progressions which may prove to be

[1] On the large sieve, *Mathematika*, **12,** 201–225 (1965).

a useful supplement to Bombieri's theorem. No account is given of other sieve methods, since these will form the theme of a later volume in this series by Professors Halberstam and Richert.[2]

H.D.

[2] This book subsequently appeared as *Sive Methods*, Academic Press (London), 1974.

BIBLIOGRAPHY

The following works will be referred to by their authors' names, or by short titles.

Bohr, H., and Cramér, H. *Die neuere Entwicklung der analytischen Zahlentheorie*, Enzyklopädie der mathematischen Wissenschaften, II.3, Heft 6, Teubner, Leipzig, 1923.

Hua, L.-K. *Die Abschätzung von Exponentialsummen und ihre Anwendung in der Zahlentheorie*, Enzyklopädie der mathematischen Wissenschaften, I.2, Heft 13, Teil 1, Teubner, Leipzig, 1959.

Ingham, A. E. *The distribution of prime numbers*, Cambridge Mathematical Tracts No. 30, Cambridge, 1932.

Landau, E. *Handbuch der Lehre von der Verteilung der Primzahlen*, 2nd ed. with an appendix by P. T. Bateman, Chelsea, New York, 1953.

Landau, E. *Vorlesungen über Zahlentheorie*, 3 vol., Hirzel, Leipzig, 1927.

Prachar, K. *Primzahlverteilung*, Springer, Berlin, 1957.

Titchmarsh, E. C. *The theory of the Riemann zeta-function*, Clarendon Press, Oxford, 1951.

NOTATION

We write $f(x) = O(g(x))$, or equivalently $f(x) \ll g(x)$, when there is a constant C such that $|f(x)| \leq Cg(x)$ for all values of x under consideration. We write $f(x) \sim g(x)$ when $\lim f(x)/g(x) = 1$ as x tends to some limit, and $f(x) = o(g(x))$ when $\lim f(x)/g(x) = 0$. Moreover, we say that $f(x) = \Omega(g(x))$ to indicate that $\limsup |f(x)|/g(x) > 0$, while $f(x) = \Omega_{\pm}(g(x))$ means that $\limsup f(x)/g(x) > 0$ and $\liminf f(x)/g(x) < 0$.

If ξ is a vector, then $\|\xi\|$ denotes its norm, while, if θ is a real number then $\|\theta\|$ denotes the distance from θ to the nearest integer. In certain contexts (see p. 32), we let $[x]$ denote the largest integer not exceeding the real number x, and we let (x) be the fractional part of x, $(x) = x - [x]$. Generally s denotes a complex variable, $s = \sigma + it$, while $\rho = \beta + i\gamma$ denotes the generic non-trivial zero of the zeta function or of a Dirichlet L function. When no confusion arises, we let γ stand for Euler's constant.

The arithmetic functions $d(n)$, $\Lambda(n)$, $\mu(n)$, and $\phi(n)$ are defined as usual. Other symbols are defined on the following pages.

1

PRIMES IN ARITHMETIC PROGRESSION

Analytic number theory may be said to begin with the work of Dirichlet, and in particular with Dirichlet's memoir of 1837 on the existence of primes in a given arithmetic progression.

Long before the time of Dirichlet it had been asserted that every arithmetic progression

$$a, a + q, a + 2q, \ldots,$$

in which a and q have no common factor, includes infinitely many primes. Legendre, who had based some of his demonstrations on this proposition, attempted to give a proof but failed. The first proof was that of Dirichlet in the memoir I have referred to (Dirichlet's *Werke*, I, pp. 313–342), and strictly speaking this proof was complete only in the case when q is a prime. For the general case, Dirichlet had to assume his class number formula, which he proved in a paper of 1839–1840 (*Werke*, I, pp. 411–496). Dirichlet states at the end of the earlier paper that originally he had a different proof, by indirect and complicated arguments, of the vital result that was needed [the fact that $L(1, \chi) \neq 0$ for each real nonprincipal character χ; see §4], but I do not think that there is any indication anywhere of its nature.

I shall follow Dirichlet's example in treating first the simpler case in which q is a prime. We can suppose that $q > 2$, for when $q = 2$ the arithmetic progression contains all sufficiently large odd numbers, and the proposition is then a triviality.

Dirichlet's starting point, as he himself says, was Euler's proof of the existence of infinitely many primes. If we write

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s},$$

for a real variable $s > 1$, then Euler's identity is

$$\zeta(s) = \prod_{p} (1 - p^{-s})^{-1}$$

for $s > 1$, where p runs through all the primes; this identity is an analytic equivalent for the proposition that every natural number can be factorized into prime powers in one and only one way. It follows from the identity that

$$\log \zeta(s) = \sum_p \sum_{m=1}^{\infty} m^{-1} p^{-ms}.$$

Since $\zeta(s) \to \infty$ as $s \to 1$ from the right, and since

$$\sum_p \sum_{m=2}^{\infty} m^{-1} p^{-ms} < \sum_p \sum_{m=2}^{\infty} p^{-m} = \sum_p \frac{1}{p(p-1)} < 1,$$

it follows that

$$\sum_p p^{-s} \to \infty$$

as $s \to 1$ from the right. This proves the existence of an infinity of primes, and proves further that the series Σp^{-1}, extended over the primes, diverges. Dirichlet's aim was to prove the analogous statements when the primes p are limited to those which satisfy the condition $p \equiv a \pmod q$.

To this end he introduced the arithmetic functions called Dirichlet's characters. Each of these is a function of the integer variable n, which is periodic with period q and is also multiplicative (without any restriction). Moreover, these functions are such that a suitable linear combination of them will produce the function which is 1 if $n \equiv a \pmod q$ and 0 otherwise.

The construction of these functions is based on the existence of a primitive root to the (prime) modulus q, or in other words on the cyclic structure of the residue classes modulo q under multiplication, when 0 is excluded. Let $\nu(n)$ denote the index of n relative to a fixed primitive root g, that is, the exponent ν for which $g^\nu \equiv n$. Let ω be a real or complex number satisfying

$$\omega^{q-1} = 1.$$

Then the typical Dirichlet character for the modulus q is

$$\omega^{\nu(n)},$$

which is uniquely defined, since the value of $\nu(n)$ is indeterminate only to the extent of the addition of a multiple of $q - 1$. The definition presupposes that n is not divisible by q, but it is convenient to complete the definition by taking the function to be 0 when n is divisible by q. There is one function for each choice of ω, and different

choices of ω give different functions; thus there are $q - 1$ such functions. Each is a periodic function of n with period q, and is multiplicative because, if

$$n \equiv n_1 n_2 \pmod q,$$

then

$$\nu(n) \equiv \nu(n_1) + \nu(n_2) \pmod{q - 1}.$$

(We have supposed here that neither n_1 nor n_2 is divisible by q, but the multiplicative property is a triviality if either of them is.)

We recall the well-known fact that $\Sigma_\omega \omega^k$ has the value $q - 1$ if k is divisible by $q - 1$ and has the value 0 otherwise. Hence

$$\sum_\omega \omega^{-\nu(a)}\omega^{\nu(n)} = \begin{cases} q - 1 & \text{if } n \equiv a \pmod q, \\ 0 & \text{otherwise,} \end{cases}$$

since $\nu(n) \equiv \nu(a) \pmod{q - 1}$ if and only if $n \equiv a \pmod q$. The expression on the left, after division by $q - 1$, is the linear combination of the various functions $\omega(n)$ that was referred to above; it serves to select from all integers n those that are congruent to the given number a to the modulus q.

For each of the possible choices for ω, Dirichlet introduced the function

$$L_\omega(s) = \sum_{\substack{n=1 \\ n \not\equiv 0 \pmod q}}^{\infty} \omega^{\nu(n)} n^{-s}$$

of the real variable s, for $s > 1$. Since the coefficient of n^{-s} is a multiplicative function of n, we have the analog of Euler's identity:

$$L_\omega(s) = \prod_{p \neq q} (1 - \omega^{\nu(p)} p^{-s})^{-1},$$

for $s > 1$. A detailed proof is easily given, on the same lines as for Euler's original identity, by considering first the finite product over $p \leq N$ and then making $N \to \infty$.

None of the factors on the right vanishes, since $|\omega^{\nu(p)} p^{-s}| = p^{-s} < \frac{1}{2}$ for $s > 1$, and as the product is absolutely convergent it follows that $L_\omega(s) \neq 0$ for $s > 1$. Taking the logarithm of both sides, we get

$$\log L_\omega(s) = \sum_{p \neq q} \sum_{m=1}^{\infty} m^{-1} \omega^{\nu(p^m)} p^{-ms}.$$

The logarithm on the left is, in principle, multivalued if ω is complex, but the value which is provided by the series on the right is obviously

the natural one to use, since it is a continuous function of s for $s > 1$ and tends to 0 as $s \to \infty$, corresponding to the fact that $L_\omega(s) \to 1$ (1 being the first term in its defining series).

Multiplying the last equation by $\omega^{-\nu(a)}$ and summing over all the values of ω, we obtain

$$(1) \qquad \frac{1}{q-1}\sum_{\omega} \omega^{-\nu(a)} \log L_\omega(s) = \sum_{p} \sum_{\substack{m=1 \\ p^m \equiv a \pmod q}}^{\infty} m^{-1} p^{-ms}.$$

The sum of all those terms on the right for which $m > 1$ is at most 1, since they are a subset of the terms considered earlier in connection with $\log \zeta(s)$. Hence the right side of (1) is

$$\sum_{p \equiv a \pmod q} p^{-s} + O(1).$$

The essential idea of Dirichlet's memoir is to prove that the left side of (1) tends to $+\infty$ as $s \to 1$. This will imply that there are infinitely many primes $p \equiv a \pmod q$, and further that the series Σp^{-1} extended over these primes is divergent.

One of the terms in the sum on the left of (1) comes from $\omega = 1$, and is simply $\log L_1(s)$. The function $L_1(s)$ is related in a simple way to $\zeta(s)$, for we have

$$L_1(s) = \sum_{\substack{n=1 \\ q \nmid n}}^{\infty} n^{-s} = (1 - q^{-s})\zeta(s).$$

Hence $L_1(s) \to +\infty$ as $s \to 1$ from the right, and therefore the same is true of $\log L_1(s)$. Hence to complete the proof it will suffice to show that, for each choice of ω other than 1,

$$\log L_\omega(s)$$

is bounded as $s \to 1$ from the right.

At this point it clarifies the situation if we observe that, provided $\omega \neq 1$, the series which defines $L_\omega(s)$, namely

$$L_\omega(s) = \sum_{\substack{n=1 \\ n \not\equiv 0 \pmod q}}^{\infty} \omega^{\nu(n)} n^{-s},$$

is convergent not only for $s > 1$ but for $s > 0$. It is, in fact, a series of the type covered by Dirichlet's test for convergence, since (a) n^{-s} decreases as n increases and has the limit 0, and (b) the sum of any

number of the coefficients $\omega^{\nu(n)}$ is bounded. The justification for (b) lies in the fact that $\omega^{\nu(n)}$ is periodic with period q, and

$$\sum_{n=1}^{q-1} \omega^{\nu(n)} = \sum_{m=0}^{q-2} \omega^{m} = 0,$$

since the index $\nu(n)$ runs through a complete set of residues to the modulus $q - 1$.

It follows further from Dirichlet's test that the series is uniformly convergent with respect to s for $s \geq \delta > 0$, and consequently $L_\omega(s)$ is a continuous function of s for $s > 0$. So to prove that $\log L_\omega(s)$ is bounded as $s \to 1$ from the right is equivalent to proving that

$$L_\omega(1) \neq 0. \tag{2}$$

Dirichlet's proof of this takes entirely different forms according as ω is real or complex. The only real value of ω is -1, since $\omega \neq 1$ now.

Suppose first that ω is complex. If we take $a = 1$, and so $\nu(a) = 0$, in (1), we get

$$\frac{1}{q-1} \sum_{\omega} \log L_\omega(s) = \sum_{\substack{p \\ p^m \equiv 1 \pmod q}} \sum_{m=1}^{\infty} m^{-1} p^{-ms}.$$

Since the terms on the right (if there are any) are positive, it follows that

$$\sum_{\omega} \log L_\omega(s) \geq 0,$$

which implies that

$$\prod_{\omega} L_\omega(s) \geq 1. \tag{3}$$

All this, of course, is for $s > 1$.

If there is some complex ω for which $L_\omega(1) = 0$, then $L_{\bar{\omega}}(1) = 0$ also, where $\bar{\omega}$ denotes the complex conjugate of ω. Thus two of the factors on the left of (3) will have the limit 0 as $s \to 1$ from the right. One other factor, namely $L_1(s)$, has the limit $+\infty$. Any other factors are certainly bounded, being continuous functions of s for $s > 0$. On examining in more detail the behavior as $s \to 1$ of the three factors mentioned, we shall get a contradiction to (3), in that the two factors with limit 0 will more than cancel the one factor with limit $+\infty$.

As regards $L_1(s)$, we have

$$L_1(s) = (1 - q^{-s})\zeta(s) < (1 - q^{-2})\zeta(s)$$

for $1 < s < 2$, and

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} < 1 + \int_1^{\infty} x^{-s}\,dx = \frac{s}{s-1}.$$

Hence

$$L_1(s) < \frac{A}{s-1}$$

for $1 < s < 2$, where A is independent of s.

As regards $L_\omega(s)$, the supposition that $L_\omega(1) = 0$ implies that, for $s > 1$,

$$L_\omega(s) = L_\omega(s) - L_\omega(1) = (s-1)L'_\omega(s_1),$$

where s_1 is some number between 1 and s. The series for $L'_\omega(s)$, namely

$$L'_\omega(s) = -\sum_{\substack{n=1 \\ n \not\equiv 0 \,(\mathrm{mod}\, q)}}^{\infty} \omega^{\nu(n)}(\log n)n^{-s},$$

is convergent for $s > 0$ by Dirichlet's test, since the function $(\log n)n^{-s}$ decreases for sufficiently large n and has the limit 0. Moreover the convergence is again uniform for $s \geq \delta > 0$, so that $L'_\omega(s)$ is continuous for $s > 0$. In particular, $|L'_\omega(s)|$ is bounded for $s > 1$, and therefore

$$|L_\omega(s)| < A_1(s-1)$$

for $s > 1$, where A_1 is independent of s. Naturally the same inequality holds with $\bar{\omega}$ in place of ω.

On using these inequalities in (3), and making $s \to 1$, we get the desired contradiction.

The argument could have been expressed more briefly by using the elements of complex function theory. As we shall see later, $L_1(s)$ has a simple pole at $s = 1$, and the supposition that $L_\omega(s)$ and $L_{\bar{\omega}}(s)$ have zeros at $s = 1$ implies that the product on the left of (3) has a zero at $s = 1$, which contradicts the inequality.

Suppose now that $\omega = -1$. The above argument is inapplicable, since the supposition that $L_{-1}(1) = 0$ would produce only a single factor with a zero at $s = 1$.

We now have

$$\omega^{\nu(n)} = (-1)^{\nu(n)} = \left(\frac{n}{q}\right),$$

and the convention made earlier that $\omega^{\nu(n)}$ is to be replaced by 0 when $n \equiv 0 \pmod q$ is in agreement with the usual convention for the Legendre symbol on the right. From now on, we abbreviate $L_{-1}(s)$ to $L(s)$, since this will be the only function with which we shall be concerned. Thus we have

$$L(s) = \sum_{n=1}^{\infty} \left(\frac{n}{q}\right) n^{-s}.$$

The aim is to prove that $L(1) \neq 0$. We already know that $L(1) \geq 0$, from the continuity of $L(s)$ and the fact that $L(s) > 0$ for $s > 1$ (by the Euler product formula). It may be worth remarking that the need to prove that $L(1) \neq 0$ is almost inevitable in the approach we are using. If it were possible for $L(1)$ to vanish, it would follow, on considering $\log L(1)$, that

$$\sum_{p} \left(\frac{p}{q}\right) p^{-s} \to -\infty \qquad \text{as } s \to 1.$$

This would imply a great preponderance of primes in those residue classes $a \pmod q$ for which a is a quadratic nonresidue, and this preponderance might (on the face of things) be such that Σp^{-1} summed over the primes in the other residue classes was convergent.

Dirichlet's proof, in the case now under consideration, is based on a relationship (which goes back to the work of Gauss) between the quadratic character $(n|q)$ and the complex exponential function $e^{2\pi i n/q}$, which I shall abbreviate to $e(n/q)$ or to $e_q(n)$. (Instead of speaking about the complex exponential function, we could speak about the qth roots of unity; but it is necessary for some purposes to be able to distinguish between one qth root of unity and another, and the complex exponential function offers the simplest way of doing this.)

Let $G(n)$ denote the so-called Gaussian sum, defined by

$$G(n) = \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) e_q(mn). \tag{4}$$

By changing the variable of summation from m to m', where $m' \equiv mn \pmod q$, we obtain the relation

$$G(n) = \left(\frac{n}{q}\right) G(1) = \left(\frac{n}{q}\right) G, \tag{5}$$

say. The argument presupposes that $n \not\equiv 0 \pmod q$, but the relation holds in the excluded case also, because then $G(n) = 0$.

Assuming that $G \neq 0$ (an assumption that will be justified later), we have

$$\left(\frac{n}{q}\right) = \frac{1}{G}\sum_{m=1}^{q-1}\left(\frac{m}{q}\right) e_q(mn).$$

Substituting this in the series for $L(1)$, we obtain

$$L(1) = \frac{1}{G}\sum_{m=1}^{q-1}\left(\frac{m}{q}\right)\sum_{n=1}^{\infty}\frac{1}{n} e_q(mn).$$

The sum of the inner series is easily deduced from that of the logarithmic series. If $|z| \leq 1$ and $z \neq 1$, we have

$$-\log(1-z) = \sum_{n=1}^{\infty}\frac{1}{n} z^n,$$

where the logarithm has its principal value. That means, in the present context, that $\arg(1-z)$ lies between $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$, since the real part of $1-z$ is positive. Taking $z = e^{i\theta}$, where $0 < \theta < 2\pi$, we have $\arg(1-z) = \frac{1}{2}(\theta - \pi)$, as is easily seen either from a picture or by calculation. Also $|1-z| = 2\sin\frac{1}{2}\theta$. Hence

$$\sum_{n=1}^{\infty}\frac{1}{n} e^{in\theta} = -\log(2\sin\tfrac{1}{2}\theta) - \tfrac{1}{2}(\theta - \pi)i.$$

Putting $\theta = 2\pi m/q$ and substituting in the formula for $L(1)$, we get

$$(6) \qquad L(1) = -\frac{1}{G}\sum_{m=1}^{q-1}\left(\frac{m}{q}\right)\left[\log\left(2\sin\frac{\pi m}{q}\right) + i\left(\frac{\pi m}{q} - \frac{\pi}{2}\right)\right].$$

As we shall prove later, the value of G is $q^{\frac{1}{2}}$ if $q \equiv 1 \pmod 4$ and $iq^{\frac{1}{2}}$ if $q \equiv 3 \pmod 4$. This compels one to distinguish two cases.

Suppose $q \equiv 3 \pmod 4$. Since $L(1)$ is real, we must have

$$(7) \qquad L(1) = -\frac{\pi}{q^{\frac{3}{2}}}\sum_{m=1}^{q-1} m\left(\frac{m}{q}\right);$$

and in fact the vanishing of the other part of the sum is easily verified by taking together the terms m and $q - m$. The last formula gives an elegant expression for $L(1)$ by a finite sum, the value of which is easily computed in any particular case. For example, if $q = 23$, we have

$$\begin{aligned}\sum_{m=1}^{q-1} m\left(\frac{m}{q}\right) &= 1+2+3+4-5+6-7+8+9-10-11+12 \\ &\quad +13-14-15+16-17+18-19-20-21-22 \\ &= -69,\end{aligned}$$

and

$$L(1) = 3\pi/23^{\frac{1}{2}}.$$

The finite sum occurring above is always an *odd* integer, for it has the same parity as

$$\sum_{m=1}^{q-1} m = \tfrac{1}{2}(q-1)q,$$

and both $\frac{1}{2}(q-1)$ and q are odd. It therefore cannot vanish, and this gives the proof of the desired result that $L(1) \neq 0$ in the case now under consideration. It is a remarkable fact that no one has yet given a simple and direct proof that the value of the finite sum in (7) is negative, though we know that this must be so from the fact that $L(s) > 0$ for $s > 1$ and consequently $L(1) \geq 0$.

Dirichlet gave another expression for $L(1)$ as a finite sum, in addition to (7), which is of great interest and which is more convenient for computation. By Euler's product formula [or alternatively from the original definition of $L(s)$], we have

$$L(s) = \left[1 - \frac{1}{2^s}\left(\frac{2}{q}\right)\right]^{-1} \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \left(\frac{n}{q}\right) n^{-s}.$$

Proceeding as before, with $s = 1$, and using the fact that

$$\mathfrak{I} \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \frac{1}{n} e_q(mn) = \begin{cases} \frac{1}{4}\pi & \text{if } 0 < m < \frac{1}{2}q, \\ -\frac{1}{4}\pi & \text{if } \frac{1}{2}q < m < q \end{cases}$$

(which is easily deduced from the sum of the logarithmic series), we obtain

$$\text{(8)} \qquad L(1) = \left[1 - \frac{1}{2}\left(\frac{2}{q}\right)\right]^{-1} \frac{1}{iq^{\frac{1}{2}}}\left(\frac{i\pi}{4}\right)\left[\sum_{m<\frac{1}{2}q}\left(\frac{m}{q}\right) - \sum_{m>\frac{1}{2}q}\left(\frac{m}{q}\right)\right]$$

$$= \frac{\pi}{[2-(2/q)]q^{\frac{1}{2}}} \sum_{m<\frac{1}{2}q}\left(\frac{m}{q}\right).$$

This shows that, for $q \equiv 3 \pmod 4$, there are always more quadratic residues than nonresidues in the first half of the range from 0 to q. Again, no direct proof is known.

Suppose $q \equiv 1 \pmod 4$. Then $G = q^{\frac{1}{2}}$, and (6) gives

$$\text{(9)} \qquad L(1) = -\frac{1}{q^{\frac{1}{2}}} \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) \log\left(2 \sin \frac{\pi m}{q}\right).$$

This can be written as

$$L(1) = \frac{\log Q}{q^{\frac{1}{2}}},$$

where

$$Q = \frac{\prod \sin(\pi N/q)}{\prod \sin(\pi R/q)},$$

in which we use R to denote the typical quadratic residue (mod q) between 0 and q, and N to denote the typical quadratic nonresidue.

To prove that $L(1) \neq 0$ is equivalent to proving that $Q \neq 1$. For this, Dirichlet had recourse to a result that had been proved by Gauss in his work on cyclotomy. This is that, for an indeterminate x,

$$\prod_R [x - e_q(R)] = \tfrac{1}{2}[Y(x) - q^{\frac{1}{2}} Z(x)]$$

and

$$\prod_N [x - e_q(N)] = \tfrac{1}{2}[Y(x) + q^{\frac{1}{2}} Z(x)],$$

where $Y(x)$ and $Z(x)$ are polynomials with integral coefficients. Assuming this, we have the identity

$$\tfrac{1}{4}[Y^2(x) - qZ^2(x)] = \prod_{m=1}^{q-1} [x - e_q(m)] = x^{q-1} + x^{q-2} + \cdots + 1.$$

Putting $x = 1$, we obtain integers $Y = Y(1)$ and $Z = Z(1)$ which satisfy the Diophantine equation

$$Y^2 - qZ^2 = 4q.$$

[Obviously Y must be divisible by q, and the present argument provides a method for solving the "negative" Pellian equation $qY_1^2 - Z^2 = 4$, when q is a prime congruent to 1 (mod 4).] We note that $Z \neq 0$, since $4q$ is not a perfect square.

The quotient Q which occurred above is expressible in terms of Y and Z, as follows. We have

$$\begin{aligned}\prod_R [1 - e_q(R)] &= \prod_R e_q(\tfrac{1}{2}R)(-2i \sin \pi R/q)\\ &= (-2i)^{(q-1)/2} e_q(\tfrac{1}{2}\textstyle\sum R) \prod_R \sin \pi R/q,\end{aligned}$$

and a similar relation with N in place of R. Now

$$\textstyle\sum R = \sum N = \tfrac{1}{4}q(q-1),$$

on grouping together the numbers R and $R' = q - R$, and similarly with N. Hence

$$Q = \frac{\prod_N \sin(\pi N/q)}{\prod_R \sin(\pi R/q)} = \frac{Y + q^{\frac{1}{2}}Z}{Y - q^{\frac{1}{2}}Z}.$$

Since $Z \neq 0$, we have the desired conclusion that $Q \neq 1$.

This completes the proof of Dirichlet's theorem for a prime modulus q, subject to the proof of the value of Gauss' sum and the proof of the result on cyclotomy which we have just used. These proofs will be given in the next two sections.

I ought perhaps to add that Dirichlet derived the finite expression (6) for $L(1)$ by a somewhat different method from that which we have used above. He started from the power series

$$\sum_{n=1}^{\infty} \left(\frac{n}{q}\right) x^n = \frac{1}{1 - x^q} \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) x^m = \frac{xf(x)}{1 - x^q},$$

say, and by putting this in the formula

$$\Gamma(s)n^{-s} = \int_0^1 x^{n-1}(\log x^{-1})^{s-1}\, dx$$

he obtained

$$\Gamma(s)L(s) = -\int_0^1 \frac{f(x)}{x^q - 1}(\log x^{-1})^{s-1}\, dx.$$

On putting $s = 1$ and expressing the rational function in the integrand as a sum of partial fractions, and integrating, he obtained (6). The two methods are essentially equivalent, but the last formula written above has some independent interest in that it serves to define $L(s)$ as a regular function of the complex variable s for all s.

2

GAUSS' SUM

In this section we evaluate the sum

$$G = \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) e_q(m),$$

where q is a prime other than 2. It is easy to prove that $G^2 = q$ if $q \equiv 1 \pmod 4$ and $G^2 = -q$ if $q \equiv 3 \pmod 4$, though this does not determine G completely. The computation is as follows. We have

$$G^2 = \sum_{m_1=1}^{q-1} \sum_{m_2=1}^{q-1} \left(\frac{m_1 m_2}{q}\right) e_q(m_1 + m_2).$$

On changing the variable of summation in the inner sum from m_2 to n, where $m_2 \equiv m_1 n \pmod q$, we get

$$G^2 = \sum_{m_1=1}^{q-1} \sum_{n=1}^{q-1} \left(\frac{n}{q}\right) e_q(m_1 + m_1 n).$$

Now we interchange the order of summation, and note that

$$\sum_{m_1=1}^{q-1} e_q[m_1(n+1)] = \begin{cases} q-1 & \text{if } n \equiv -1 \pmod q, \\ -1 & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} G^2 &= q\left(\frac{-1}{q}\right) + \sum_{n=1}^{q-1} \left(\frac{n}{q}\right)(-1) \\ &= q\left(\frac{-1}{q}\right) = \begin{cases} q & \text{if } q \equiv 1 \pmod 4, \\ -q & \text{if } q \equiv 3 \pmod 4, \end{cases} \end{aligned}$$

as stated above.

The sign of G was determined by Gauss only after many and varied unsuccessful attempts.[1] Since then several proofs have been

[1] See Gauss, *Werke* II, p. 156.

given, based on a variety of different methods.[2] As Gauss himself remarked, any proof of the exact value of G must take account of the particular ordering of the qth roots of unity, which is provided by the complex exponential function. If instead of G we consider the sum

$$\sum_{m=1}^{q-1} \left(\frac{m}{q}\right) \zeta^m,$$

where ζ is any qth root of unity (other than 1), its sign cannot be specified, for it follows from (5) of the preceding section that if ζ is replaced by ζ^n, where n is a quadratic nonresidue (mod q), the sign of the sum gets reversed. The evaluation of G^2 given above would, however, apply equally well to the sum with an unspecified ζ.

The method used by Dirichlet in 1835 (*Werke* I, pp. 237–256) to evaluate G is probably the most satisfactory of all that are known. It is based on Poisson's summation formula, and it has the advantage that once the proof has been embarked upon, no special ingenuity is called for.

It is first necessary to express the definition of G in a form which does not contain explicitly the symbol of quadratic character. With the same meaning for R and N as in §1, we have

$$G = \sum_R e_q(R) - \sum_N e_q(N) = 1 + 2\sum_R e_q(R).$$

This can be written equivalently as

$$G = \sum_{x=0}^{q-1} e_q(x^2),$$

for x^2 assumes the value 0 once and assumes each value R twice.

Dirichlet's method actually evaluates the more general sum

$$S = \sum_{n=0}^{N-1} e^{2\pi i n^2/N},$$

where N is any positive integer, and the answer is that

$$S = \begin{cases} (1+i)N^{\frac{1}{2}} & \text{if } N \equiv 0 \pmod 4, \\ N^{\frac{1}{2}} & \text{if } N \equiv 1 \pmod 4, \\ 0 & \text{if } N \equiv 2 \pmod 4, \\ iN^{\frac{1}{2}} & \text{if } N \equiv 3 \pmod 4. \end{cases}$$

[2] See Landau, *Vorlesungen* I, pp. 157–171, and Estermann, *J. London Math. Soc.*, **20,** 66–67 (1945).

Here $N^{\frac{1}{2}}$ denotes the positive square root. It may be as well to add also that i denotes the same square root of -1 as occurs in $e^{2\pi i n^2/N}$.

Poisson's summation formula states that, under certain conditions on the function $f(x)$,

$$\sum_{n=A}^{B}{}' f(n) = \sum_{\nu=-\infty}^{\infty} \int_A^B f(x) e^{2\pi i \nu x}\, dx,$$

where Σ' means that the end terms of the sum, corresponding to $n = A$ and $n = B$, are to be replaced by $\frac{1}{2}f(A)$ and $\frac{1}{2}f(B)$. In the series on the right it may be necessary to take the terms ν and $-\nu$ together to ensure convergence, but actually it will not be necessary in the present application. Sufficient conditions for the validity of the formula are that $f(x)$ is a real function which is continuous and monotonic in stretches. From the point of view of analysis these are severe restrictions, but they are quite adequate for most applications.

Poisson's summation formula, which is an extremely useful tool in analytic number theory, is easily deduced (under the above restrictions) from the basic theorem concerning Fourier series, which was first rigorously proved by Dirichlet himself in 1829 (*Werke* I, pp. 117–132). Let $f_1(x)$ coincide with $f(x)$ for $0 \leq x < 1$ and be defined elsewhere by periodicity with period 1; then $f_1(x)$ is continuous for $0 < x < 1$ but has (in general) an ordinary discontinuity at $x = 0$, its values on the left and on the right being $f(1)$ and $f(0)$ respectively. The Fourier series of $f_1(x)$ is

$$\tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos 2\pi\nu x + b_\nu \sin 2\pi\nu x),$$

where the coefficients are given by Fourier's formulas:

$$\tfrac{1}{2}a_\nu = \int_0^1 f(x)\cos 2\pi\nu x\, dx, \qquad \tfrac{1}{2}b_\nu = \int_0^1 f(x)\sin 2\pi\nu x\, dx.$$

The theorem in question is that the above series converges for all x, and that its sum is $f_1(x)$ at a point of continuity of $f_1(x)$, and is the mean of the left and right values of $f_1(x)$ at a point of ordinary discontinuity. Thus, taking $x = 0$, we get

$$\tfrac{1}{2}[f(0) + f(1)] = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} a_\nu = \sum_{\nu=-\infty}^{\infty} \int_0^1 f(x)\cos 2\pi\nu x\, dx.$$

This is the case $A = 0$, $B = 1$ of Poisson's summation formula, and the general case follows on replacing $f(x)$ by $f(x + n)$, for $n = A, A + 1, \ldots, B - 1$, and adding the results.

For the application to Gauss' sum, we take $f(x) = \cos 2\pi x^2/N$ and $f(x) = \sin 2\pi x^2/N$ and combine the results. Thus

$$\begin{aligned} S &= \sum_{\nu=-\infty}^{\infty} \int_0^N e^{2\pi i\nu x + 2\pi i x^2/N}\, dx \\ &= N \sum_{\nu=-\infty}^{\infty} \int_0^1 e^{2\pi i N(x^2+\nu x)}\, dx \\ &= N \sum_{\nu=-\infty}^{\infty} e^{-\frac{1}{2}\pi i N\nu^2} \int_{\frac{1}{2}\nu}^{1+\frac{1}{2}\nu} e^{2\pi i N y^2}\, dy, \end{aligned}$$

where in the last step we have put $x + \frac{1}{2}\nu = y$. The value of $e^{-\frac{1}{2}\pi i N\nu^2}$ is 1 if ν is even, and is $e^{-\frac{1}{2}\pi i N} = i^{-N}$ if ν is odd. We therefore divide the sum over ν into two parts, according as ν is even or odd, and we put $\nu = 2\mu$ or $2\mu - 1$ as the case may be. This gives

$$S = N \sum_{\mu=-\infty}^{\infty} \int_{\mu}^{\mu+1} e^{2\pi i N y^2}\, dy + N i^{-N} \sum_{\mu=-\infty}^{\infty} \int_{\mu-\frac{1}{2}}^{\mu+\frac{1}{2}} e^{2\pi i N y^2}\, dy.$$

Each series of integrals fits together to give

$$\int_{-\infty}^{\infty} e^{2\pi i N y^2}\, dy.$$

This is a convergent integral, and it is a matter of indifference whether we construe it in the narrow sense, as

$$\lim_{Y\to\infty} \int_{-Y}^{Y},$$

or in the wider sense as

$$\lim_{Y,Z\to\infty} \int_{-Y}^{Z}.$$

For if $Y' > Y > 0$ we have

$$\int_Y^{Y'} e^{2\pi i N y^2}\, dy = \tfrac{1}{2}\int_{Y^2}^{Y'^2} z^{-\frac{1}{2}} e^{2\pi i N z}\, dz,$$

and by the second mean value theorem, or by integration by parts, this has absolute value $O(Y^{-1})$ as $Y \to \infty$. The convergence of the integral in the wider sense justifies our earlier remark that, in the present application of Poisson's summation formula, it is not necessary to take together the terms ν and $-\nu$.

Resuming the evaluation of S, we have

$$S = N(1 + i^{-N})\int_{-\infty}^{\infty} e^{2\pi i N y^2}\, dy,$$

and this implies, on putting $y = N^{-\frac{1}{2}}u$, that

$$S = (1 + i^{-N})CN^{\frac{1}{2}},$$

where C is an absolute constant. This constant is most easily evaluated by putting $N = 1$, whereupon $S = 1$ and we get $C = (1 + i^{-1})^{-1}$. Hence

$$S = \frac{1 + i^{-N}}{1 + i^{-1}} N^{\frac{1}{2}},$$

and this gives the four values stated earlier, according to the residue class to which N belongs to the modulus 4.

3

CYCLOTOMY

Cyclotomy is concerned with the properties of the roots of unity of a given order, with particular reference to their algebraic character.[1] Our first object must be to establish the result quoted in §1, and this we can do without going very deeply into the theory. Afterward I shall digress briefly from the main theme of these lectures to discuss two topics in cyclotomy which are of general interest.

We shall be concerned only with roots of unity of prime order. Let q be a prime other than 2, and let ζ be a qth root of unity other than 1. Then the entire set of qth roots of unity other than 1 is

(1) $$\zeta, \zeta^2, \zeta^3, \ldots, \zeta^{q-1},$$

and the sum of these numbers is -1. By using this relation, together with the relation $\zeta^q = 1$, we can express any polynomial in ζ with integral coefficients in the form

$$a_1\zeta + a_2\zeta^2 + \cdots + a_{q-1}\zeta^{q-1},$$

where $a_1, \ldots, a_{q-1}$ are integers. Moreover the expression in this form is unique, since the cyclotomic polynomial

$$x^{q-1} + x^{q-2} + \cdots + 1,$$

of which ζ is a zero, is irreducible over the rational field, and therefore ζ cannot satisfy an equation of lower degree with integral coefficients.

Let g be a primitive root to the modulus q, and let $\nu(n)$ denote the index of n relative to g. As n assumes the values $1, 2, \ldots, q-1$, its index $\nu(n)$ assumes the same values in another order.

Now consider any factorization of $q - 1$, say

$$q - 1 = ef.$$

[1] The standard references are Bachmann's *Kreisteilung* of 1872 and Mathews's *Theory of Numbers* of 1892.

The roots of unity ζ^n enumerated in (1) can be subdivided according to the residue class to which $v(n)$ belongs to the modulus e. There will be e such sets, each comprising f numbers. The sums of the various subsets are called the Gaussian periods of f terms, and are denoted by $\eta_0, \eta_1, \ldots, \eta_{e-1}$. Thus

$$\eta_j = \sum_{\substack{n=1 \\ v(n) \equiv j \pmod e}}^{q-1} \zeta^n.$$

It is obviously not essential to restrict j to the range $0 \leq j < e$; the last equation can be used for all j, and then η_j is periodic in j with period e.

In particular, if $e = 2$ and $f = \frac{1}{2}(q-1)$, we get the two periods

$$\eta_0 = \sum \zeta^R, \qquad \eta_1 = \sum \zeta^N,$$

where, as earlier, R and N run through the quadratic residues and nonresidues respectively, in the range from 1 to $q-1$. If we fix $\zeta = e^{2\pi i/q}$, we can deduce the values of these two periods from the value of Gauss' sum. For then

$$\eta_0 - \eta_1 = \sum_{m=1}^{q-1} \left(\frac{m}{q}\right) e_q(m) = G = \varepsilon q^{\frac{1}{2}},$$

where ε is 1 or i according as $q \equiv 1$ or 3 (mod 4). Since $\eta_0 + \eta_1 = -1$, it follows that

$$\eta_0 = \tfrac{1}{2}(-1 + \varepsilon q^{\frac{1}{2}}), \qquad \eta_1 = \tfrac{1}{2}(-1 - \varepsilon q^{\frac{1}{2}}).$$

In the general case, if we choose $\zeta = e^{2\pi i/q}$, the value of η_0 is uniquely determined, and in fact

$$\eta_0 = e^{-1} \sum_{x=1}^{q-1} e_q(x^e).$$

But the individual values of $\eta_1, \ldots, \eta_{e-1}$ will depend on the choice of the primitive root, and they may get permuted if this is replaced by another primitive root.

Now let $F(\zeta)$ be any polynomial in ζ, say

$$F(\zeta) = \sum_{r=1}^{q-1} A_r \zeta^r,$$

and suppose $F(\zeta)$ has the property that

$$F(\zeta^m) = F(\zeta)$$

whenever $\nu(m) \equiv 0 \pmod{e}$. Then, by the uniqueness of representation of a polynomial, we have

$$A_r = A_s$$

whenever $r \equiv sm \pmod{q-1}$, and this holds for all m with $\nu(m) \equiv 0 \pmod{e}$. Hence A_r depends only on the residue class $\pmod{e}$ to which $\nu(r)$ belongs. Grouping together the terms in the same residue class, we obtain

$$F(\zeta) = A_1\eta_1 + \cdots + A_e\eta_e.$$

Thus $F(\zeta)$ is a linear combination of the Gaussian periods.

We have tacitly supposed that the coefficients A_r are integers (or rational numbers), and it is only under some such restriction that we can appeal to the uniqueness of representation of a polynomial in ζ. But the result holds equally if the coefficients A_r are themselves polynomials in an indeterminate x with integral coefficients, for then it holds for every integral value of x, and therefore identically in x.

We apply this, with $e = 2$, to the polynomial

$$F(\zeta) = \prod_R (x - \zeta^R).$$

When written in the standard form, the coefficients $A_r(x)$ are polynomials in x with integral coefficients. If m is any integer with $\nu(m) \equiv 0 \pmod{2}$, then m is a quadratic residue, and

$$F(\zeta^m) = \prod_R (x - \zeta^{Rm}) = \prod_R (x - \zeta^R) = F(\zeta).$$

Hence $F(\zeta)$ has the property postulated above, and it follows that

$$F(\zeta) = A_0(x)\eta_0 + A_1(x)\eta_1.$$

[Actually $A_0(x) = A_2(x)$, in the notation we have just been using.]

Substituting the values of η_0 and η_1, we obtain

$$\begin{aligned}\prod_R (x - \zeta^R) &= \tfrac{1}{2}A_0(x)(-1 + \varepsilon q^{\frac{1}{2}}) + \tfrac{1}{2}A_1(x)(-1 - \varepsilon q^{\frac{1}{2}}) \\ &= \tfrac{1}{2}[Y(x) - \varepsilon q^{\frac{1}{2}} Z(x)],\end{aligned}$$

where $Y(x)$, $Z(x)$ are polynomials with integral coefficients. If we replace ζ by ζ^g, then ζ^R becomes ζ^N, where N is a typical quadratic nonresidue, and η_0 and η_1 become interchanged. This has the effect of changing ε into $-\varepsilon$. Hence

$$\prod_N (x - \zeta^N) = \tfrac{1}{2}[Y(x) + \varepsilon q^{\frac{1}{2}} Z(x)].$$

This proves the result quoted in §1; it was used there in the case $q \equiv 1 \pmod 4$, and so with $\varepsilon = 1$.

I shall now discuss two topics connected with cyclotomy, for the sake of their intrinsic interest. They are (a) Gauss' theorem on the roots of unity of order q, when q is a prime of the form $2^k + 1$, and (b) Kummer's problem on the cubic periods.

GAUSS' THEOREM

Gauss' theorem asserts that if q is a prime of the form $2^k + 1$ (e.g., if q is 3 or 5 or 17 or 257 or 65537), each qth root of unity can be expressed in terms of rational numbers by using a succession of square root signs. From this assertion, with a few supplementary observations, one deduces that, for the values of q in question, it is possible to inscribe a regular polygon of q sides in a given circle by a Euclidean construction using ruler and compasses only.

We consider the various choices of e and f that are possible:

$$e_1 = 2, \qquad f_1 = \tfrac{1}{2}(q-1);$$
$$e_2 = 4, \qquad f_2 = \tfrac{1}{4}(q-1);$$
$$\cdots$$
$$e_k = 2^k, \qquad f_k = \frac{1}{2^k}(q-1) = 1.$$

For the choice e_r, there are e_r Gaussian periods of f_r terms, which we shall denote by

$$\eta_1^{(r)}, \ldots, \eta_e^{(r)} \qquad (e = e_r)$$

to indicate the dependence on r.

We have already evaluated the two periods $\eta_1^{(1)}, \eta_2^{(1)}$, and they are $\frac{1}{2}(-1 \pm \sqrt{q'})$, where $q' = q$ if $q \equiv 1 \pmod 4$ and $q' = -q$ if $q \equiv 3 \pmod 4$. The latter cannot happen if $q > 3$.

Now consider the four periods $\eta_1^{(2)}, \eta_2^{(2)}, \eta_3^{(2)}, \eta_4^{(2)}$. By definition,

$$\eta_j^{(2)} = \sum_{\nu(n) \equiv j \pmod 4} \zeta^n.$$

The expression

$$(x - \eta_1^{(2)})(x - \eta_3^{(2)}),$$

considered as a polynomial in ζ, is unaltered if we replace ζ by ζ^m, provided $\nu(m) \equiv 0 \pmod 2$, for the effect of this is either to leave $\eta_1^{(2)}$ and $\eta_3^{(2)}$ unchanged or to interchange them. Hence, by our earlier result,

$$(x - \eta_1^{(2)})(x - \eta_3^{(2)}) = A_1(x)\eta_1^{(1)} + A_2(x)\eta_2^{(1)},$$

where $A_1(x)$, $A_2(x)$ have integral coefficients. It follows that the coefficients of the quadratic in x on the left are expressible by rational numbers and $\sqrt{q}$. Hence $\eta_1^{(2)}$, $\eta_3^{(2)}$ are expressible by means of two square root signs, and similarly for $\eta_2^{(2)}$, $\eta_4^{(2)}$.

The argument continues; at the next step, the eight periods fall into the four groups:

$$\eta_1^{(3)},\ \eta_5^{(3)};\qquad \eta_2^{(3)},\ \eta_6^{(3)};\qquad \eta_3^{(3)},\ \eta_7^{(3)};\qquad \eta_4^{(3)},\ \eta_8^{(3)};$$

and the two in each group can be evaluated in terms of the four periods $\eta_j^{(2)}$ by use of another square root sign.

Finally, we come to the 2^k periods of one term; these are just $\zeta, \zeta^2, \ldots, \zeta^{q-1}$. Thus each of these is expressible by means of rational numbers and k square root signs. The k ambiguities of sign attaching to the square roots give the $2^k (= q - 1)$ roots of unity.

This proves Gauss' theorem in its first form. For the inscription of a regular polygon of q sides in a circle, it suffices to have the number $\cos 2\pi/q$, which determines the first point of sub-division of the circle. Now

$$2 \cos 2\pi/q = \zeta + \zeta^{-1},$$

and this is one of the periods of two terms which arise at the penultimate stage of the preceding construction, for then $e = 2^{k-1} = \frac{1}{2}(q - 1)$, and the exponents 1 and -1 on the right above are just the values of n for which the index $\nu(n)$ is divisible by e. Thus we can construct the length $2\cos 2\pi/q$ from a unit length by solving a succession of quadratic equations. But in order that this construction shall be capable of realization geometrically, it is necessary that all the quadratic equations shall have real roots. Thus we need to know that all the periods $\eta_j^{(r)}$, with $r \leq k - 1$, are real. This is in fact the case. For if η is one such period, then $\bar{\eta}$ is obtained from η by changing ζ into ζ^{-1}, and this has the effect of replacing $\nu(n)$ by $\nu(-n)$. Now $g^{2^{k-1}} = g^{\frac{1}{2}(q-1)} \equiv -1 \pmod{q}$, and therefore $\nu(-1) = 2^{k-1}$, and so is divisible by e for each of the values $e = 2, 4, \ldots, 2^{k-1}$. Hence the condition of summation $\nu(n) \equiv j \pmod{e}$ is unaltered if $\nu(n)$ is replaced by $\nu(-n)$, and therefore $\eta = \bar{\eta}$, that is, η is real.

KUMMER'S PROBLEM

Kummer's problem relates to the three periods of $\frac{1}{3}(q - 1)$ terms that exist when $q - 1$ is a multiple of 3. These are

$$\eta_0 = \sum \zeta^A, \qquad \eta_1 = \sum \zeta^B, \qquad \eta_2 = \sum \zeta^C,$$

where A runs through those numbers n of the set $1, 2, \ldots, q-1$ whose indices are divisible by 3, and B through those whose indices are $\equiv 1 \pmod 3$, and C through those whose indices are $\equiv 2 \pmod 3$. The numbers A constitute the cubic residues $(\operatorname{mod} q)$, and the numbers B and C constitute the two classes of cubic nonresidues.

If we choose $\zeta = e^{2\pi i/q}$, as we shall do henceforward, the value of η_0 is uniquely determined, and in fact

$$1 + 3\eta_0 = \sum_{x=0}^{q-1} e_q(x^3),$$

since the function x^3 assumes the value 0 once and assumes each of the values A three times. But the values of η_1 and η_2 cannot be distinguished from one another unless we specify also the primitive root by which the index is defined (and, as far as I know, there is no simple and general way of doing so).

The values of η_0, η_1, η_2 can be expressed in terms of a Gaussian sum which is similar to the sum G defined in (4) of §1, but is formed with a cubic character instead of a quadratic character. Let ω be a complex cube root of unity, and define

$$\chi(n) = \omega^{\nu(n)}$$

for $n \not\equiv 0 \pmod q$, and put $\chi(n) = 0$ for $n \equiv 0 \pmod q$. Define

$$\tau = \sum_{n=1}^{q-1} \chi(n)\, e_q(n).$$

We first prove that $|\tau| = q^{\frac{1}{2}}$. We have

$$|\tau|^2 = \sum_{n_1=1}^{q-1} \sum_{n_2=1}^{q-1} \chi(n_1)\bar{\chi}(n_2) e_q(n_1 - n_2),$$

and, with a computation similar to that at the beginning of §2, this is

$$\sum_{n_1=1}^{q-1} \sum_{n=1}^{q-1} \bar{\chi}(n) e_q(n_1 - nn_1) = q\bar{\chi}(1) + \sum_{n=1}^{q-1} \bar{\chi}(n)(-1) = q.$$

This proves the assertion; and we can now write

$$\tau = q^{\frac{1}{2}} e^{i\theta},$$

where $\theta = \theta(q)$. θ is uniquely determined, as an angle, except for sign, for the only ambiguity in the definition of τ lies in the possibility of replacing χ by $\bar{\chi}$, and this has the effect of changing τ into $\bar{\tau}$ [since $\chi(-1) = 1$ for a cubic character] and so of changing θ into $-\theta$.

The expressions for η_0, η_1, η_2 in terms of τ are very easily derived. It is convenient to put

$$1 + 3\eta_j = z_j \qquad (j = 0, 1, 2).$$

Then

$$z_0 = \sum_{x=0}^{q-1} [1 + \chi(x) + \bar{\chi}(x)]e_q(x) = \tau + \bar{\tau} = 2q^{\frac{1}{2}} \cos \theta.$$

Similarly

$$\begin{aligned} z_1 &= \sum_{x=0}^{q-1} [1 + \omega^2\chi(x) + \omega\bar{\chi}(x)]e_q(x) \\ &= \omega^2\tau + \omega\bar{\tau} = 2q^{\frac{1}{2}} \cos\left(\theta - \frac{2\pi}{3}\right), \end{aligned}$$

and again

$$\begin{aligned} z_2 &= \sum_{x=0}^{q-1} [1 + \omega\chi(x) + \omega^2\bar{\chi}(x)]e_q(x) \\ &= \omega\tau + \omega^2\bar{\tau} = 2q^{\frac{1}{2}} \cos\left(\theta + \frac{2\pi}{3}\right). \end{aligned}$$

Kummer's problem is essentially that of determining the distribution of the angle $\theta = \theta(q)$, or rather of $\cos \theta$, as q runs through the primes. But he put the problem in a more specific form. The three numbers z_0, z_1, z_2 are the roots of a cubic equation with integral coefficients, since this is true of the three periods η_0, η_1, η_2. It follows from the above expressions in terms of θ that there is just one of the three numbers z_0, z_1, z_2 in each of the intervals

$$(-2\sqrt{q}, -\sqrt{q}), \qquad (-\sqrt{q}, \sqrt{q}), \qquad (\sqrt{q}, 2\sqrt{q}).$$

Kummer asked: With what frequencies does the number z_0, which is uniquely defined for each q, fall in each of these intervals? On somewhat limited numerical evidence he conjectured, very tentatively, that the relative frequencies may be in the ratios 1:2:3, but more extensive computation by Mrs. Lehmer[2] made this appear unlikely. Recently, Heath-Brown and Patterson[3] have shown that the θ are uniformly distributed, so that the limiting ratios are 1:1:1. In their proof they use, among other things, the techniques which we develop in §24.

It may be of interest to show that $\cos 3\theta$ can be expressed in terms of the representation of q in the form

$$4q = a^2 + 27b^2;$$

[2] *Math. Tables and Aids to Computation*, **10**, 194–202 (1956).
[3] *J. reine angew. Math.*, **310**, 111–130 (1979).

it is easily proved that this representation is unique, except of course for the signs of a and b. We have (with variables of summation running from 1 to $q - 1$)

$$\begin{aligned}\tau^2 &= \sum_x \sum_y \chi(x)\chi(y)e_q(x + y) \\ &= \sum_x \sum_t \chi^2(x)\chi(t)e_q[x(1 + t)] \\ &= \sum_{t \not\equiv -1} \chi(t) \sum_x \bar{\chi}(x)e_q[x(1 + t)] \\ &= \sum_{t \not\equiv -1} \chi(t)\chi(1 + t)\bar{\tau}.\end{aligned}$$

Multiplying by τ, we get

$$\tau^3 = q \sum_t \chi[t(1 + t)] = q(A + B\omega),$$

where A and B are integers. Obviously

$$\bar{\tau}^3 = q(A + B\bar{\omega}),$$

and on multiplying the two equations together we get

$$q = (A + B\omega)(A + B\bar{\omega}) = A^2 - AB + B^2,$$

or

$$4q = (2A - B)^2 + 3B^2.$$

We now prove that B is divisible by 3, this being necessary because without this stipulation the representation of q in the above form is not unique. We observe that τ is an algebraic integer, and that

$$\tau^3 = \left[\sum_x \chi(x)e_q(x)\right]^3 = \sum_x \chi^3(x)e_q(3x) + 3\xi,$$

where ξ is an algebraic integer. Subtracting from this the corresponding equation for $\bar{\tau}^3$, we see that $\tau^3 - \bar{\tau}^3$ is divisible by 3. But

$$\tau^3 - \bar{\tau}^3 = qB(\omega - \bar{\omega}) = iqB\sqrt{3};$$

hence the rational integer B is divisible by 3.

We now have

$$4q = a^2 + 27b^2,$$

where

$$a = 2A - B, \qquad b = \tfrac{1}{3}B;$$

and

$$q^{\frac{3}{2}} e^{3i\theta} = \tau^3 = q(A + B\omega) = \tfrac{1}{2}q(a + iB\sqrt{3}).$$

Hence

$$\cos 3\theta = \frac{a}{2\sqrt{q}}.$$

This determines $\cos 3\theta$ except for sign. But the ambiguous sign, arising from the unknown sign of a, can also be specified, for it can be shown that

$$a \equiv 1 \pmod 3.$$

To prove this, we consider the number N of solutions of the congruence

$$v^3 \equiv u(u + 1) \qquad (\text{mod } q).$$

For $u \equiv 0$ or -1 there is just one value of v, and for any other u there are either three values of v or none. Hence $N \equiv 2 \pmod 3$. On the other hand,

$$N = \sum_{u=0}^{q-1} \{1 + \chi[u(u + 1)] + \chi^2[u(u + 1)]\}$$

$$= q + (A + B\omega) + (A + B\omega^2) = q + a.$$

Hence $a \equiv 2 - q \equiv 1 \pmod 3$.

In conclusion we remark that the cubic equation mentioned earlier, whose roots are z_1, z_2, z_3, is simply

$$z^3 - 3qz - qa = 0.$$

This follows easily from the expressions for the z_j in terms of τ and $\bar{\tau}$. We have

$$\sum z_j = (\tau + \bar{\tau}) + (\omega^2\tau + \omega\bar{\tau}) + (\omega\tau + \omega^2\bar{\tau}) = 0,$$

$$\sum z_j^2 = (\tau + \bar{\tau})^2 + (\omega^2\tau + \omega\bar{\tau})^2 + (\omega\tau + \omega^2\bar{\tau})^2$$

$$= 6\tau\bar{\tau} = 6q,$$

and

$$z_1 z_2 z_3 = \tau^3 + \bar{\tau}^3 = q(2A - B) = qa.$$

For further information on Kummer's problem, see Mathews, *Theory of Numbers*, §§ 196 and 197; and Hasse, *Vorlesungen über Zahlentheorie* (2nd ed., 1964), §20.6.

4

PRIMES IN ARITHMETIC PROGRESSION: THE GENERAL MODULUS

Dirichlet's proof of the existence of primes in a given arithmetic progression, in the general case when the modulus q is not necessarily a prime, is in principle a natural extension of that in the special case. But the proof given in §1 that $L_{\omega}(1) \neq 0$ when $\omega = -1$, which involved separate consideration of the cases $q \equiv 1$ and $q \equiv 3 \pmod 4$, does not extend to give the analogous result that is needed when q is composite.

We now suppose that q is any positive integer other than 1. (We do not exclude $q = 2$, as we did in §1, though it will in fact be a trivial case.)

The functions that take the place of the functions $\omega^{\nu(n)}$, where $\omega^{q-1} = 1$, are Dirichlet's characters to the modulus q. These are functions of an integer variable n which are periodic with period q and multiplicative without restriction. The typical function is denoted by $\chi(n)$; it is defined initially when n is relatively prime to q, but the definition is then conveniently extended by defining $\chi(n)$ to be 0 when $(n, q) > 1$. The number of these functions will be $\phi(q)$.

Dirichlet's characters to a given modulus can be regarded as a particular case of the characters of an Abelian group, the group in question here being that of the relatively prime residue classes (mod q) combined by multiplication. But I shall follow Dirichlet in giving a direct and constructive account of them. This is partly for historical reasons, in that Dirichlet's work preceded by several decades the development of group theory, and partly for a mathematical reason, namely that the group in question has a simple and interesting structure which is obscured if one treats it as one treats the general Abelian group.

Consider first the case when q is a power of a prime other than 2, say $q = p^{\alpha}$. Here the construction of §1 extends quite naturally. There is a primitive root, and the theory of the index applies, the only difference being that the modulus to which the index is defined is now $\phi(p^{\alpha}) = p^{\alpha-1}(p - 1)$ in place of $p - 1$. We define the characters to the modulus p^{α} by taking any real or complex number ω

which satisfies

$$\omega^{\phi(p^\alpha)} = 1,$$

and putting

$$\chi(n) = \omega^{\nu(n)} \qquad \text{for } (n, p) = 1,$$

where $\nu(n)$ denotes the index of n relative to a fixed primitive root of the modulus p^α. The number of characters is $\phi(p^\alpha)$.

The position is a little more complicated when $q = 2^\alpha$. If $\alpha = 1$ there is only one relatively prime residue class, and only one character, the value of which is always 1 (for odd n, of course). If $\alpha = 2$, so that $\phi(2^\alpha) = 2$, there is a primitive root, namely -1, and the preceding construction applies: the characters are $\omega^{\nu(n)}$, where $\omega^2 = 1$. The effect is to give two characters, one of which is always 1 and the other of which is 1 or -1 according as $n \equiv 1$ or $-1 \pmod 4$. But if $\alpha \geq 3$ there is no primitive root (mod 2^α); as a substitute for this we have the fact that every relatively prime residue class is representable uniquely as

$$(-1)^\nu 5^{\nu'},$$

where ν is defined to the modulus 2 and ν' is defined to the modulus $\frac{1}{2}\phi(2^\alpha) = 2^{\alpha-2}$. By analogy with the previous construction, we define the characters in the present case by

$$\chi(n) = \omega^\nu(\omega')^{\nu'},$$

where

$$\omega^2 = 1 \qquad \text{and} \qquad (\omega')^{2^{\alpha-2}} = 1.$$

The number of characters is $2^{\alpha-1} = \phi(2^\alpha)$.

In the general case, when

$$q = 2^\alpha p_1^{\alpha_1} p_2^{\alpha_2} \dots,$$

we define the characters to the modulus q as products of arbitrary characters to the various prime power moduli. If $\chi(n; 2^\alpha)$ denotes any character to the modulus 2^α, and similarly for the other prime powers, the general character to the modulus q is given by

$$\chi(n) = \chi(n; 2^\alpha)\chi(n; p_1^{\alpha_1})\chi(n; p_2^{\alpha_2})\dots,$$

provided $(n, q) = 1$. (The last proviso could be omitted, for if n has a factor in common with q, one of the characters in the product on the right will be 0.) The total number of characters is

$$\phi(2^\alpha)\phi(p_1^{\alpha_1})\phi(p_2^{\alpha_2})\dots = \phi(q).$$

It is plain that these characters are distinct arithmetic functions (this being a consequence of the fact that each index assumes all values to its appropriate modulus), and that each function is a periodic and multiplicative function of n. One of the characters, got by taking all the ω's to be 1, has always the value 1, for $(n, q) = 1$, and this is called the principal character and denoted by χ_0.

The characters to a given modulus q form themselves a group under multiplication, with the principal character as the unit element. This group, which has $\phi(q)$ elements, is in fact isomorphic to the multiplicative group of the relatively prime residue classes $(\text{mod}\, q)$. The isomorphism is most easily demonstrated by rewriting the definition of $\chi(n)$ in terms of the complex exponential function. For the modulus p^α, we have

$$\omega = e^{2\pi i m/\phi(p^\alpha)} = e[m/\phi(p^\alpha)],$$

and the different choices of ω correspond to different choices of the integer m to the modulus $\phi(p^\alpha)$. So

$$\chi(n; p^\alpha) = e\left[\frac{mv}{\phi(p^\alpha)}\right],$$

where v is the index of n relative to a particular primitive root of p^α. In the case 2^α, we have

$$\chi(n; 2^\alpha) = e\left(\frac{mv}{2} + \frac{m'v'}{2^{\alpha-2}}\right),$$

where $n \equiv (-1)^v\, 5^{v'} \pmod{2^\alpha}$. Putting these formulas together, we get

$$\chi(n) = e\left[\frac{m_0 v_0}{2} + \frac{m_0' v_0'}{2^{\alpha-2}} + \frac{m_1 v_1}{\phi(p_1^{\alpha_1})} + \frac{m_2 v_2}{\phi(p_2^{\alpha_2})} + \cdots\right] \tag{1}$$

for $(n, q) = 1$, where $m_0, m_0', m_1, m_2 \ldots$ are integers which take all values modulo the corresponding denominators. The definition is symmetric in the m's and the v's, and we see that multiplication relative to n (with χ fixed) corresponds to addition of the vectors

$$(v_0, v_0', v_1, v_2, \ldots),$$

and that multiplication relative to χ (with n fixed) corresponds to addition of the vectors

$$(m_0, m_0', m_1, m_2, \ldots),$$

in each case with respect to the appropriate moduli. This duality renders visible the isomorphism mentioned earlier. We have assumed above, for simplicity of exposition, that the exponent

α in 2^α is at least 3. It will of course be plain that if $\alpha = 2$ the second of the two terms corresponding to 2^α is to be omitted, and that if $\alpha = 1$ both terms are to be omitted.

The characters have an important property that can be expressed in either of two equivalent forms. In the first form, it states that

$$\sum_n \chi(n) = \begin{cases} \phi(q) & \text{if } \chi = \chi_0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where the summation is over any representative set of residues (mod q), though it suffices to take a set of relatively prime residues, since $\chi(n) = 0$ for the others. The truth of the above statement is an immediate deduction from the representation of the general character in (1). For the summation over n is equivalent to a summation over $\nu_0, \nu'_0, \nu_1, \nu_2,\ldots$, each to its respective modulus, and this gives 0 unless each of $m_0, m'_0, m_1, m_2,\ldots$ is congruent to 0 with respect to its corresponding modulus. In that case, $\chi = \chi_0$, and all the values of $\chi(n)$ are 1 for n relatively prime to q, and the value of the sum is $\phi(q)$.

The second form of the property is that

$$\sum_\chi \chi(n) = \begin{cases} \phi(q) & \text{if } n \equiv 1 \pmod{q}, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where the summation is over all the $\phi(q)$ characters. The same proof applies, but with the m's and ν's interchanged; the only case in which the sum does not vanish is that in which all the ν's are 0, and then $n \equiv 1 \pmod{q}$. It may be of interest to remark that if the characters are defined axiomatically, that is, by their periodic and multiplicative properties, instead of by construction, then (2) is readily deducible from the definition but (3) is not. To prove this, one has either to use similar ideas to those we have used in the construction, or to appeal to the basis theorem for Abelian groups.

Using (3), we can prove that *any arithmetic function $X(n)$ that is multiplicative and has period q, and is* 0 *when $(n, q) > 1$ but not always* 0, *is one of the $\phi(q)$ characters $\chi(n)$*. For if $(c, q) = 1$, we have

$$\sum_n X(n)\bar{\chi}(n) = \sum_n X(cn)\bar{\chi}(cn) = X(c)\bar{\chi}(c)\sum_n X(n)\bar{\chi}(n).$$

Unless $X(c) = \chi(c)$ for all c, the sum must be 0. If this is so for each χ, then

$$0 = \sum_\chi \chi(m)\sum_n X(n)\bar{\chi}(n) = \sum_n X(n)\sum_\chi \chi(m)\bar{\chi}(n) = \phi(q)X(m).$$

This gives $X(m) = 0$ for all m, contrary to hypothesis.

We can also use (3) to construct, as in §1, a linear combination of the characters which selects those integers n which fall in a given residue class (mod q). If $(a, q) = 1$, then

$$\frac{1}{\phi(q)}\sum_{\chi}\bar{\chi}(a)\chi(n) = \begin{cases} 1 & \text{if } n \equiv a \pmod{q}, \\ 0 & \text{otherwise}; \end{cases} \tag{4}$$

for we have $\bar{\chi}(a)\chi(n) = \chi(n')$, and $n' \equiv 1 \pmod{q}$ if and only if $n \equiv a \pmod{q}$.

The L functions for a general modulus q are defined, in the first place for $s > 1$, by

$$L(s, \chi) = \sum_{n=1}^{\infty} \chi(n)n^{-s}.$$

As in §1, each of them has an Euler product expression:

$$L(s, \chi) = \prod_{p} [1 - \chi(p)p^{-s}]^{-1},$$

and $L(s, \chi) \neq 0$ for $s > 1$. We have

$$\log L(s, \chi) = \sum_{p}\sum_{m=1}^{\infty} m^{-1}\chi(p^m)p^{-ms},$$

and on forming a linear combination of these logarithms and using the relation (4), we obtain

$$\frac{1}{\phi(q)}\sum_{\chi}\bar{\chi}(a)\log L(s, \chi) = \sum_{\substack{p \\ p^m \equiv a \pmod{q}}}\sum_{m=1}^{\infty} m^{-1}p^{-ms}. \tag{5}$$

As in §1, the right side is

$$\sum_{p \equiv a \pmod{q}} p^{-s} + O(1)$$

as $s \to 1$ from the right. Thus our object, as before, is to prove that the left side of (5) tends to $+\infty$ as $s \to 1$ from the right.

The term corresponding to the principal character χ_0 is

$$\frac{1}{\phi(q)}\log L(s, \chi_0).$$

By the Euler product formula, we have

$$L(s, \chi_0) = \zeta(s)\prod_{p|q}(1 - p^{-s}), \tag{6}$$

and therefore $\log L(s, \chi_0) \to +\infty$ as $s \to 1$. It therefore suffices to prove that, for $\chi \neq \chi_0$, $\log L(s, \chi)$ is bounded as $s \to 1$, and again this is equivalent to proving that $L(1, \chi) \neq 0$.

If χ is a complex character, that is, a character whose values are not all real, so that $\bar{\chi} \neq \chi$, this follows as in §1 from the inequality

$$\prod_{\chi} |L(s, \chi)| \geq 1 \qquad \text{for } s > 1,$$

which is proved in the same way as before by taking $a = 1$ in (5).

It remains to prove that $L(1, \chi) \neq 0$ when χ is a real character other than the principal character. Dirichlet deduced this from his famous class-number formula, an account of which will be given in §6. But to complete the proof of Dirichlet's theorem now, I shall deviate from the historical order of discovery and give a simple proof due to de la Vallée Poussin,[1] which is based on complex function theory.

For this proof we need to know a little about the behavior of the L functions as functions of a complex variable s. We write $s = \sigma + it$, as is customary in this subject. The series which defines $L(s, \chi)$ is absolutely convergent for $\sigma > 1$, and is uniformly convergent with respect to s for $\sigma \geq 1 + \delta$ for any positive δ. Hence the L functions are defined for $\sigma > 1$ and are regular functions of s there. We can, however, easily prove that each of them can be continued analytically so as to be regular for $\sigma > 0$, except that $L(s, \chi_0)$ has a simple pole at $s = 1$.

We deal first with $L(s, \chi_0)$, and in view of the simple relation between $L(s, \chi_0)$ and $\zeta(s)$ given in (6), it will suffice to consider $\zeta(s)$. We transform the definition $\zeta(s) = \Sigma n^{-s}$, which is applicable for $\sigma > 1$, into a form that is applicable more generally for $\sigma > 0$. This is done by partial summation, but it is a technical convenience to use integrals rather than sums. We have

$$\begin{aligned}\zeta(s) = \sum_{n=1}^{\infty} n^{-s} &= \sum_{n=1}^{\infty} n[n^{-s} - (n+1)^{-s}] \\ &= s \sum_{n=1}^{\infty} n \int_{n}^{n+1} x^{-s-1}\, dx \\ &= s \int_{1}^{\infty} [x] x^{-s-1}\, dx.\end{aligned}$$

We now put $[x] = x - (x)$, so that (x) denotes the fractional part of x. This gives

$$\zeta(s) = \frac{s}{s-1} - s \int_{1}^{\infty} (x) x^{-s-1}\, dx. \tag{7}$$

[1] "Recherches analytiques sur la théorie des nombres premiers." Deuxième partie. *Ann. Soc. Sci. Bruxelles*, **20**, 281–362 (1896).

The integral on the right is absolutely convergent for $\sigma > 0$, and uniformly for $\sigma \geq \delta > 0$, and so represents a regular function of s for $\sigma > 0$. Thus $\zeta(s)$ is meromorphic for $\sigma > 0$, its only pole being a simple pole at $s = 1$ with residue 1. In view of (6), the same is true of $L(s, \chi_0)$, except that the residue is

$$\prod_{p|q} (1 - p^{-1}) = q^{-1}\phi(q).$$

There is a similar calculation when $\chi \neq \chi_0$. If we define temporarily

$$S(x) = \sum_{n \leq x} \chi(n),$$

then

$$L(s, \chi) = \sum_{n=1}^{\infty} \chi(n)n^{-s} = \sum_{n=1}^{\infty} S(n)[n^{-s} - (n+1)^{-s}]$$

(8)
$$= s\int_1^{\infty} S(x)x^{-s-1}\,dx,$$

for $\sigma > 1$. Since $\chi \neq \chi_0$, it follows from (2) that $\Sigma\chi(n)$ over any q consecutive integers is 0, and therefore that $S(x)$ is a bounded function of x. Thus the last integral gives the analytic continuation of $L(s, \chi)$ as a regular function for $\sigma > 0$.

Suppose now that χ is a real nonprincipal character (mod q) and that $L(1, \chi) = 0$. Then $L(s, \chi)$ has a zero at $s = 1$, and the product

$$L(s, \chi)L(s, \chi_0)$$

is regular at $s = 1$ and therefore regular for $\sigma > 0$. Since $L(2s, \chi_0)$ is regular and different from 0 for $\sigma > \frac{1}{2}$, the function

$$\psi(s) = \frac{L(s, \chi)L(s, \chi_0)}{L(2s, \chi_0)}$$

is regular for $\sigma > \frac{1}{2}$. We observe further that $\psi(s) \to 0$ as $s \to \frac{1}{2}$ from the right, since $L(2s, \chi_0) \to +\infty$.

The Euler product formula for $\psi(s)$ contains only factors corresponding to primes that do not divide q, and indeed contains only factors corresponding to primes for which $\chi(p) = 1$, since if $\chi(p) = -1$ the factor is

$$\frac{(1 + p^{-s})^{-1}(1 - p^{-s})^{-1}}{(1 - p^{-2s})^{-1}} = 1.$$

Thus we get

$$\psi(s) = \prod_{\chi(p)=1} \left(\frac{1 + p^{-s}}{1 - p^{-s}}\right).$$

This holds for $\sigma > 1$. If there were no primes with $\chi(p) = 1$ we should have $\psi(s) = 1$ for all $\sigma > 1$, and therefore by analytic continuation for all $\sigma > \frac{1}{2}$, and this is contrary to the fact that $\psi(s) \to 0$ as $s \to \frac{1}{2}$.

Plainly the above product can be written as a Dirichlet series:

$$\psi(s) = \sum_{n=1}^{\infty} a_n n^{-s},$$

where $a_n \geq 0$ and $a_1 = 1$. This series is only valid, however, for $\sigma > 1$ (as far as we know).

Since $\psi(s)$ is regular for $\sigma > \frac{1}{2}$, it has an expansion in powers of $s - 2$ with a radius of convergence at least $\frac{3}{2}$. This power series is

$$\psi(s) = \sum_{m=0}^{\infty} \frac{1}{m!} \psi^{(m)}(2)(s-2)^m.$$

We can calculate $\psi^{(m)}(2)$ from the Dirichlet series, and we obtain

$$\psi^{(m)}(2) = (-1)^m \sum_{n=1}^{\infty} a_n (\log n)^m n^{-2} = (-1)^m b_m,$$

say, where $b_m \geq 0$. Hence

$$\psi(s) = \sum_{m=0}^{\infty} \frac{1}{m!} b_m (2-s)^m,$$

and this holds for $|2 - s| < \frac{3}{2}$. If $\frac{1}{2} < s < 2$, then since all the terms are nonnegative we have

$$\psi(s) \geq \psi(2) \geq 1,$$

and this contradicts the fact that $\psi(s) \to 0$ as $s \to \frac{1}{2}$. Thus the hypothesis that $L(1, \chi) = 0$ is disproved.[2]

We have therefore completed the proof of Dirichlet's theorem that *there are infinitely many primes* $p \equiv a \pmod q$, *and the series* Σp^{-1} *summed over such primes is divergent.*

[2] A somewhat different proof, but on similar general lines, was given by Landau in 1905 (see, for example, Prachar, Chap. 4, Satz 4.2). There is also an elementary but rather complicated proof due to Mertens, which will be found in Landau, *Vorlesungen* I, Satz 152.

5

PRIMITIVE CHARACTERS

Many results about characters and L functions take a simple form only for the so-called primitive characters, though they may be capable of extension, with complications, to imprimitive characters. We shall now explain the distinction between these two types of character, and afterward investigate in detail the real primitive characters.

Let $\chi(n)$ be any character to the modulus q other than the principal character. If $(n, q) > 1$, then $\chi(n) = 0$; if $(n, q) = 1$, then $\chi(n) \neq 0$, being a root of unity, and is a periodic function of n with period q. It is possible, however, that for values of n restricted by the condition $(n, q) = 1$, the function $\chi(n)$ may have a period less than q. If so, we say that χ is *imprimitive*, and otherwise *primitive*.[1] It is a matter of personal preference whether one includes the principal character among the imprimitive characters; I prefer to leave it unclassified.

Let $\chi(n)$ be a nonprincipal character to the modulus q which is imprimitive, and let q_1 be its least period. Then $q_1 < q$; and $q_1 > 1$, for otherwise we should have $\chi(n) = \chi(1) = 1$ for all n satisfying $(n, q) = 1$, contrary to the supposition that χ is not the principal character. Further, q_1 is a factor of q, for by a familiar argument if q and q_1 are periods then so is (q, q_1), and therefore this number cannot be less than q_1.

We shall prove that $\chi(n)$ is identical, *when* $(n, q) = 1$, with a character $\chi_1(n)$ to the modulus q_1; but before we can prove this we must define $\chi_1(n)$. Of course, we define $\chi_1(n)$ to be $\chi(n)$ if $(n, q) = 1$; and if $(n, q_1) = 1$ but $(n, q) > 1$, we choose any integer t such that

(1) $$(n + tq_1, q) = 1$$

and define $\chi_1(n) = \chi(n + tq_1)$. Such an integer exists, for it suffices to have

$$(n + tq_1, r) = 1,$$

[1] Alternative terms are *improper* and *proper*.

where[2] r is the product of those prime power constituents of q which are relatively prime to q_1. The choice of t, subject to (1), is immaterial, since the value of $\chi(n + tq_1)$ will be the same.

We have now defined $\chi_1(n)$ when $(n, q_1) = 1$, and of course we take it to be 0 when $(n, q_1) > 1$. Plainly $\chi_1(n)$ is periodic with period q_1, and its multiplicative property follows easily from that of $\chi(n)$. Further, $\chi_1(n)$ is not always 0 when $(n, q_1) = 1$, for $\chi_1(1) = \chi(1) = 1$. Hence, by a result proved in §4, it is one of the $\phi(q_1)$ characters to the modulus q_1. The values of $\chi_1(n)$ when $(n, q_1) = 1$ include the values of $\chi(n)$ when $(n, q) = 1$, and so cannot be periodic with period less than q_1; nor can they all be 1. Hence $\chi_1(n)$ is a primitive character to the modulus q_1. We have now proved that *to an imprimitive character* χ (mod q) *there corresponds a proper factor* q_1 *of* q *and a primitive character* χ_1 (mod q_1) *such that*

$$\chi(n) = \begin{cases} \chi_1(n) & \textit{if } (n, q) = 1, \\ 0 & \textit{if } (n, q) > 1. \end{cases} \tag{2}$$

We say that χ_1 induces χ. It is clear that if q_1 and χ_1 are given, and q is any proper multiple of q_1, the above definition of χ does in fact produce a character (mod q).

For example, the Legendre symbol $(n|p)$ is an imprimitive character (mod p^α) if $\alpha > 1$, being induced by the same character (mod p); but this is a particularly simple case, since here the conditions $(n, q) = 1$ and $(n, q_1) = 1$ are synonymous. Or again the Legendre symbol $(n|p_1)$ induces an imprimitive character to the modulus $p_1 p_2$ (where $p_2 \neq p_1$) by the definition

$$\chi(n) = \begin{cases} \left(\dfrac{n}{p_1}\right) & \text{if } (n, p_1 p_2) = 1, \\ 0 & \text{if } (n, p_1 p_2) > 1. \end{cases}$$

As we saw in §4, any character (mod q) is representable as

$$\chi(n) = \chi(n; p_1^{\alpha_1})\chi(n; p_2^{\alpha_2})\ldots,$$

where $q = p_1^{\alpha_1} p_2^{\alpha_2}\ldots$, and the characters on the right are to the moduli indicated. (We allow p_1 to be 2 here.) It is easily seen that χ is primitive if and only if each of the characters on the right is primitive. If χ is imprimitive, one or more of the characters on the right is either principal or imprimitive, and in the latter case $\chi(n; p^\alpha) = \chi(n; p^\beta)$, where $1 \leq \beta < \alpha$. Then q_1 is the product of the prime powers $p_i^{\beta_i}$, and χ_1 is the product of the characters $\chi(n; p_i^{\beta_i})$, but omitting any factors that are principal characters.

[2] r is not the same as $q/(q, q_1)$.

Expressed in terms of the representation by the complex exponential function, in (1) of §4, a character is primitive if and only if all the m_i are relatively prime to the corresponding p_i (with an obvious modification for m_0 and m_0' depending on whether $\alpha > 2$ or $\alpha = 2$).

The relation (2) between an imprimitive character χ and the primitive character χ_1 which induces it implies a simple relation between the corresponding L functions. By the Euler product formula,

$$(3) \qquad \begin{aligned} L(s,\chi) &= \prod_{p\nmid q}[1-\chi(p)p^{-s}]^{-1} \\ &= \prod_{p\nmid q}[1-\chi_1(p)p^{-s}]^{-1} \\ &= L(s,\chi_1)\prod_{p\mid q}[1-\chi_1(p)p^{-s}]. \end{aligned}$$

The above argument is valid only for $\sigma > 1$, where the infinite products converge; but by analytic continuation the result remains true for $\sigma > 0$, and indeed in the whole s plane, as we shall see later. In particular, $L(1,\chi_1) \neq 0$ implies $L(1,\chi) \neq 0$.

We now turn to the real primitive characters, which are of particular interest in several ways. The obvious question is: For what moduli does there exist a real primitive character (or possibly more than one), and how can such characters be expressed in terms of quadratic residue symbols? The general nature of the answer is that only for certain types of q does a real primitive character exist, and it is then expressible (for $n > 0$) as

$$\chi(n) = \left(\frac{d}{n}\right),$$

where the symbol on the right is Kronecker's extension of Legendre's symbol, and $d = \pm q$. In some cases, but not in all, d can be both $+q$ and $-q$, and then there are two characters.

We have seen that a primitive character (mod q) is a product of primitive characters with the prime power constituents of q as moduli. Consider first a prime power p^α for which $p > 2$. The character is

$$e\left[\frac{m\nu(n)}{p^{\alpha-1}(p-1)}\right], \qquad \text{for } (n,p) = 1.$$

Since $e(x)$ is real only if $x \equiv 0$ or $\frac{1}{2}$ (mod 1), and since a possible value of $\nu(n)$ is 1, this is a real function only if m is divisible by $\frac{1}{2}p^{\alpha-1}(p-1)$. We must therefore have $\alpha = 1$, for if $\alpha > 1$ we should have m divisible by p and the character would be imprimitive. We can take

$m = \frac{1}{2}(p-1)$, since $m = 0$ would give the principal character, and now the function becomes

$$e[\tfrac{1}{2}\nu(n)] = (-1)^{\nu(n)} = \left(\frac{n}{p}\right).$$

Thus p^α must be p and $\chi(n; p^\alpha)$ must be $(n|p)$.

Now consider the modulus 2^α, where $\alpha \geq 2$ of necessity, since for $\alpha = 1$ there is only the principal character. If $\alpha = 2$, there is just one nonprincipal character, namely

$$\chi_4(n) = \begin{cases} 1 & \text{if } n \equiv 1 \pmod 4, \\ -1 & \text{if } n \equiv -1 \pmod 4, \end{cases} \tag{4}$$

and this is obviously primitive. If $\alpha \geq 3$, the general character is

$$e\left(\frac{m\nu}{2} + \frac{m'\nu'}{2^{\alpha-2}}\right),$$

where $0 \leq m < 2$, $0 \leq m' < 2^{\alpha-2}$, and ν, ν' are defined by

$$n \equiv (-1)^\nu 5^{\nu'} \pmod{2^\alpha}.$$

The character can only be real if m' is divisible by $2^{\alpha-3}$, and if $\alpha > 3$ this implies that the character is imprimitive. We must have $\alpha = 3$, and there are the two possibilities $m_0 = 0$, $m_0' = 1$ and $m_0 = 1$, $m_0' = 1$. [The other possibility, $m_0 = 1$, $m_0' = 0$ leads to $\chi_4(n)$, which is imprimitive to the modulus 8.] The first of these gives a character, which we shall denote by $\chi_8(n)$, according to the rule

$$\chi_8(n) = \begin{cases} 1 & \text{if } n \equiv \pm 1 \pmod 8, \\ -1 & \text{if } n \equiv \pm 3 \pmod 8; \end{cases} \tag{5}$$

and the second possibility gives $\chi_8(n)\chi_4(n)$. Both these are primitive.

Thus the only prime power moduli to which there exist real primitive characters are:

$$\begin{cases} p\,(>2) \text{ with the character } (n|p), \\ 4 \text{ with the character } \chi_4(n), \\ 8 \text{ with the characters } \chi_8(n) \text{ and } \chi_4(n)\chi_8(n). \end{cases} \tag{6}$$

A real primitive character exists to the modulus q if and only if q is a product of such moduli, subject to the factors being relatively prime, and the character is then the product of the corresponding characters given above. There are two of them if and only if q includes the factor 8. We shall call the moduli listed above the basic moduli and the characters the basic characters.

We can express most of the basic characters, if we limit ourselves to positive values of n, in terms of Jacobi's symbol $(m|n)$, which is defined (by multiplying together the corresponding Legendre symbols) when n is odd and positive. We have[3]

$$(7)\qquad \begin{cases} \chi_4(n) = \left(\dfrac{-1}{n}\right) = \left(\dfrac{-4}{n}\right), \\ \chi_8(n) = \left(\dfrac{2}{n}\right) = \left(\dfrac{8}{n}\right), \\ \chi_4(n)\chi_8(n) = \left(\dfrac{-2}{n}\right) = \left(\dfrac{-8}{n}\right), \end{cases}$$

provided n is odd, which it naturally is when the modulus is 4 or 8. We also have, by the law of quadratic reciprocity,[4]

$$(8)\qquad \left(\frac{n}{p}\right) = \left(\frac{p'}{n}\right), \qquad \text{where } p' = (-1)^{(p-1)/2}p,$$

provided n is odd. But here the limitation to odd n is an undesirable restriction. It is removed by employing Kronecker's extension of Legendre's symbol, by which one puts

$$\left(\frac{p'}{2}\right) = \left(\frac{2}{p}\right),$$

and, more generally,

$$\left(\frac{p'}{2^{\nu}m}\right) = \left(\frac{2^{\nu}m}{p}\right).$$

With this extension, relation (8) holds whether n is odd or even. It holds also in the more general form

$$\left(\frac{n}{P}\right) = \left(\frac{P'}{n}\right), \qquad \text{where } P' = (-1)^{\frac{1}{2}(P-1)}P,$$

if $P = p_1p_2\ldots$; that is, if P is any square-free odd positive integer; for then $P' = p'_1p'_2\ldots$.

We have now expressed all the basic characters by quadratic residue symbols; they are

$$\left(\frac{-4}{n}\right), \quad \left(\frac{8}{n}\right), \quad \left(\frac{-8}{n}\right), \quad \left(\frac{p'}{n}\right),$$

[3] Landau, *Vorlesungen*, Satz 92 and Satz 93.
[4] Landau, *Vorlesungen*, Satz 95.

and the modulus of each character is the absolute value of the upper number. Moreover, we have

$$\left(\frac{d_1 d_2}{n}\right) = \left(\frac{d_1}{n}\right)\left(\frac{d_2}{n}\right)$$

provided d_1 and d_2 are relatively prime. This is a consequence of the multiplicative property of the Jacobi symbol (and so of the Legendre symbol) if n is odd, and a consequence of Kronecker's definition if n is even (which it can only be if d_1, d_2 are odd).

It follows that *the real primitive characters are identical with the symbols* $(d|n)$, *where* d *is a product of relatively prime factors of the form*

$$(9) \qquad -4, \quad 8, \quad -8, \quad (-1)^{\frac{1}{2}(p-1)}p \qquad (p > 2);$$

and the symbol is a real primitive character to the modulus $|d|$.

There is an intimate connection between the real primitive characters and the theory of binary quadratic forms, or the equivalent theory of quadratic fields. We prove, in the first place, that the numbers d described above are identical with the numbers that arise as *fundamental discriminants* in the theory of quadratic forms, or as *discriminants* in the theory of quadratic fields.

The numbers $(-1)^{\frac{1}{2}(p-1)}p$ are all congruent to 1 (mod 4), and the products of relatively prime factors (i.e., distinct factors) each of this form comprise all square-free integers, positive and negative, that are congruent to 1 (mod 4). In addition, we get all such numbers multiplied by -4, that is, all numbers $4N$, where N is square-free and congruent to 3 (mod 4). Finally, we get all such numbers multiplied by ± 8, which is equivalent to saying all numbers $4N$, where N is congruent to 2 (mod 4). Thus we get (a) all integers, positive and negative, that are $\equiv 1 \pmod 4$ and square-free, and (b) all integers, positive and negative, of the form $4N$, where $N \equiv 2$ or 3 (mod 4) and square-free.

These are just the discriminants of quadratic fields. For a quadratic field is generated by $\sqrt{N}$, where N is a square-free integer (positive or negative); and an integral basis of the field is given by

$$(1, \sqrt{N}) \qquad \text{if } N \equiv 2 \text{ or } 3 \pmod 4,$$

$$(1, \tfrac{1}{2} + \tfrac{1}{2}\sqrt{N}) \qquad \text{if } N \equiv 1 \pmod 4.$$

The discriminant, being the square of the determinant formed by an integral basis and the (algebraically) conjugate basis, is $4N$ in the first case and N in the second case. Hence the discriminants are just the numbers described in (a) and (b) above.

In the theory of quadratic forms, the discriminant of

$$ax^2 + bxy + cy^2$$

is the familiar algebraic invariant $D = b^2 - 4ac$. In this theory one presupposes that D is not a perfect square, since in that case the form has rational linear factors. Thus a discriminant is an integer, not a square, which is congruent to 0 or 1 (mod 4). A fundamental discriminant is one which has the property that all forms of that discriminant have $(a, b, c) = 1$. We can easily prove that the fundamental discriminants are just the numbers d described in (a) and (b). First, if $D = d$ and $(a, b, c) = m > 1$, then m^2 divides d, and therefore d must be of the type (b) and m must be 2. But then $a = 2a_1$, $b = 2b_1$, $c = 2c_1$, and

$$b_1^2 - 4a_1c_1 = \tfrac{1}{4}d = N,$$

which contradicts the fact that $N \equiv 2$ or $3 \pmod 4$. Second, if $D \neq d$, we easily see that $D = dm^2$ for some $m > 1$, and then there is either the imprimitive form with coefficients

$$m, m, -\tfrac{1}{4}m(d - 1)$$

or the imprimitive form with coefficients

$$m, 0, -\tfrac{1}{4}md,$$

of discriminant D. This proves the assertion.

In the theory of quadratic fields, the value of $(d|p)$ determines the way in which a prime p factorizes in the quadratic field of discriminant d; it remains a prime if $(d|p) = -1$, and factorizes into two prime ideals if $(d|p) = 1$. Similarly, in the theory of quadratic forms, p is not representable by any form of (fundamental) discriminant d if $(d|p) = -1$, but is representable by at least one form if $(d|p) = 1$.

In connection with primitive real characters, it may be noted that $\chi(-1)$ *has the value* $+1$ *or* -1 *according as d is positive or negative.* It is sufficient to prove this for the "prime discriminants" listed in (9), as the general character is a product of basic characters, and both the value of $\chi(-1)$ and the sign of d are multiplicative. For $d = -4$ the character is $\chi_4(n)$ and $\chi_4(-1) = -1$. For $d = 8$ the character is $\chi_8(n)$, and $\chi_8(-1) = 1$. For $d = -8$ the character is $\chi_4(n)\chi_8(n)$, and $\chi_4(-1)\chi_8(-1) = -1$. For $d = (-1)^{\frac{1}{2}(p-1)}p$ the character is $(n|p)$, and for $n = -1$ it is $+1$ or -1 according as $p \equiv 1$ or $-1 \pmod 4$, that is, according as d is positive or negative. Hence the result.

Thus a real primitive character is associated with a real quadratic field or with an imaginary quadratic field, according to the value of $\chi(-1)$.

Finally, we observe that the L function of any real primitive character can now be expressed as

$$L(s, \chi) = \sum_{n=1}^{\infty} \left(\frac{d}{n}\right) n^{-s} = \prod_{p} \left[1 - \left(\frac{d}{p}\right) p^{-s}\right]^{-1}$$

for $\sigma > 1$.

6

DIRICHLET'S CLASS NUMBER FORMULA

Dirichlet's class number formula, in its simplest and most striking form, was conjectured by Jacobi[1] in 1832 and (as we said in §1) proved in full by Dirichlet in 1839.

There are two stages in Dirichlet's work. In the first stage, the class number of quadratic forms of given (fundamental) discriminant d is related to the value of $L(1, \chi)$, where χ is the real primitive character $(d|n)$. This relation renders visible the fact that $L(1, \chi) > 0$. In the second stage, the value of $L(1, \chi)$ is expressed in terms of a finite sum by an argument which is essentially the same as that used in §1.

In this section we shall give the substance of Dirichlet's work, but to avoid excessive length we shall quote a number of results concerning quadratic forms from Landau's *Vorlesungen* I. We cannot follow Dirichlet in detail, because he used the notation

$$ax^2 + 2bxy + cy^2$$

for a quadratic form, whereas (following Lagrange and most modern writers) we shall use the notation

$$ax^2 + bxy + cy^2.$$

The forms of given (fundamental) discriminant d fall into classes of mutually equivalent forms under linear substitutions of the type

$$x = \alpha x' + \beta y', \qquad y = \gamma x' + \delta y', \tag{1}$$

with integral coefficients α, β, γ, δ satisfying $\alpha\delta - \beta\gamma = 1$. We call these unimodular substitutions. As Lagrange showed, every class contains at least one form whose coefficients satisfy the inequalities

$$|b| \leq |a| \leq |c|,$$

[1] See p. 51 below, and Bachmann, *Kreisteilung*, Vorlesung 20, or H. J. S. Smith, *Report on the Theory of Numbers*, §121.

and it follows easily that the number of classes, for a given discriminant d, is finite.[2]

If d is negative, the forms of discriminant d are definite. Half of them are positive definite and half are negative definite, the latter being obtained from the former by replacing a, b, c by $-a$, $-b$, $-c$. It is obviously sufficient to consider the positive definite forms, which is equivalent to saying that we restrict ourselves to forms with $a > 0$. If d is positive, each of the forms of discriminant d is indefinite. It is therefore equivalent to some form with $a > 0$, for we can choose some positive number represented properly by the form (that is, with x and y relatively prime), and any such number occurs as the first coefficient of some equivalent form. We can select a representative from each class of equivalent forms with $a > 0$, and it is convenient to do so. We denote the number of classes of forms (positive definite if $d < 0$) by $h(d)$.

There is always at least one form of discriminant d, namely, the principal form

$$\text{(2)} \qquad \begin{cases} x^2 - \frac{1}{4}dy^2 & \text{if } d \equiv 0 \pmod 4, \\ x^2 + xy - \frac{1}{4}(d-1)y^2 & \text{if } d \equiv 1 \pmod 4). \end{cases}$$

Hence $h(d)$ is a positive integer.

In the relationship between $h(d)$ and $L(1, \chi)$, the proof of which represents the first stage of Dirichlet's work, there intervenes a factor depending on the automorphs of the forms of discriminant d, that is, the unimodular substitutions that transform a form into itself. There are always two trivial automorphs, namely, the identity $x = x'$, $y = y'$ and the negative identity $x = -x'$, $y = -y'$. If $d < 0$, there are in general no others, but there are two exceptions to this: when $d = -3$ or -4. In both these cases there is only one class of forms, represented by the principal form. If $d = -3$, the principal form is $x^2 + xy + y^2$, and this has the additional automorphs

$$x = -y', \quad y = x' + y', \qquad \text{and} \qquad x = x' + y', \quad y = -x'$$

and their negatives. If $d = -4$, the principal form is $x^2 + y^2$, and this has the additional automorph

$$x = y', \qquad y = -x'$$

[2] Landau, *Vorlesungen*, Satz 197.

and its negative. We denote by w the number of automorphs, so that

$$w = \begin{cases} 2 & \text{if } d < -4, \\ 4 & \text{if } d = -4, \\ 6 & \text{if } d = -3. \end{cases} \tag{3}$$

(Another interpretation for w is that it is the number of roots of unity in the quadratic field of discriminant d.)

The position is quite different when $d > 0$. Each form has infinitely many automorphs, and these are determined by the solutions of Pell's equation

$$t^2 - du^2 = 4. \tag{4}$$

For the form with coefficients a, b, c, the automorphs are given by[3]

$$\begin{cases} \alpha = \tfrac{1}{2}(t - bu), & \beta = -cu, \\ \gamma = au, & \delta = \tfrac{1}{2}(t + bu). \end{cases} \tag{5}$$

The trivial automorphs correspond to the trivial solutions $t = \pm 2$, $u = 0$ of Pell's equation. The equation (4) has infinitely many solutions, and if t_0, u_0 is that solution with $t_0 > 0$, $u_0 > 0$ for which u_0 is least, then all solutions are given by[4]

$$\tfrac{1}{2}(t + u\sqrt{d}) = \pm[\tfrac{1}{2}(t_0 + u_0\sqrt{d})]^n, \tag{6}$$

where n is an integer (positive or negative). That (5) actually does give an automorph is easily verified by factorizing the form $ax^2 + bxy + cy^2$. We have

$$ax^2 + bxy + cy^2 = a(x - \theta y)(x - \theta' y), \tag{7}$$

where

$$\theta = \frac{-b + \sqrt{d}}{2a}, \qquad \theta' = \frac{-b - \sqrt{d}}{2a}, \tag{8}$$

and the effect of the unimodular substitution with the coefficients (5) is expressed by

$$\begin{cases} x - \theta y = \tfrac{1}{2}(t - u\sqrt{d})(x' - \theta y'), \\ x - \theta' y = \tfrac{1}{2}(t + u\sqrt{d})(x' - \theta' y'); \end{cases} \tag{9}$$

the product of the constant factors is 1 by (4).

[3] Landau, *Vorlesungen*, Satz 202.
[4] Landau, *Vorlesungen*, Satz 111.

We now turn to the question of the total number of representations of a positive integer n by a representative set of forms of given (fundamental) discriminant d. This question was answered (implicitly, at least) in the classical theory of quadratic forms, developed by Lagrange and further by Gauss.

If $d < 0$, so that the forms are positive definite, the number of representations of n by any form is finite. We denote by $R(n)$ the total number of representations by the various forms of a representative set. But if $d > 0$ there are infinitely many representations, since any one representation gives rise to an infinity of others by the application of the automorphs of the form. We shall select one representation from each such set, and call it primary, and it will transpire that the number of primary representations is finite. If x, y and X, Y are two representations of the same integer that are related by an automorph, then by (9) we have

$$\frac{x - \theta' y}{x - \theta y} = \frac{\frac{1}{2}(t + u\sqrt{d})}{\frac{1}{2}(t - u\sqrt{d})} \cdot \frac{X - \theta' Y}{X - \theta Y}.$$

Let $\varepsilon = \frac{1}{2}(t_0 + u_0\sqrt{d}) > 1$. Then, by (6),

$$\tfrac{1}{2}(t + u\sqrt{d}) = \pm\varepsilon^m, \qquad \tfrac{1}{2}(t - u\sqrt{d}) = \pm\varepsilon^{-m},$$

for some integer m. There is just one choice of m (for given X and Y) which will ensure that

$$1 \le \frac{x - \theta' y}{x - \theta y} < \varepsilon^2, \tag{10}$$

and then by choice of the ambiguous sign we can further ensure that

$$x - \theta y > 0. \tag{11}$$

A representation that satisfies these two conditions will be called primary. The number of primary representations of a given integer n by a given form is finite, since the product of the linear forms $x - \theta y$ and $x - \theta' y$ is n/a by (7), and their quotient is bounded both ways by (10). For $d > 0$ we denote by $R(n)$ the total number of primary representations of n by a representative set of forms of discriminant d.

The basic result of the theory of quadratic forms is as follows.[5]

[5] Landau, *Vorlesungen*, Satz 204.

If $n > 0$ *and* $(n, d) = 1$ *then*

$$R(n) = w \sum_{m|n} \left(\frac{d}{m}\right), \tag{12}$$

where w *is given by* (3) *if* $d < 0$ *and* $w = 1$ *if* $d > 0$.

This is proved by expressing $R(n)$ in terms of the number of solutions of the congruence $z^2 \equiv d \pmod{4n}$, and then evaluating this number in terms of quadratic character symbols.

The basic idea in the first stage of Dirichlet's work is to determine, from the above expression for $R(n)$, the average value of $R(n)$ as n varies. It is convenient (and it suffices for the purpose in view) to limit oneself to values of n that are relatively prime to d. We have

$$w^{-1} \sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n) = \sum_{\substack{m_1 m_2 \le N \\ (m_1 m_2, d) = 1}} \left(\frac{d}{m_1}\right)$$

$$= \sum_{m_1 \le \sqrt{N}} \left(\frac{d}{m_1}\right) \sum_{\substack{m_2 \le N/m_1 \\ (m_2,d)=1}} 1 + \sum_{\substack{m_2 < \sqrt{N} \\ (m_2,d)=1}} \sum_{\sqrt{N} < m_1 \le N/m_2} \left(\frac{d}{m_1}\right),$$

since the first sum comprises all pairs m_1, m_2 for which $m_1 \le \sqrt{N}$ and the second sum all pairs for which $m_1 > \sqrt{N}$. The first inner sum is

$$\frac{N}{m_1} \frac{\phi(|d|)}{|d|} + O[\phi(|d|)],$$

so the first double sum is

$$N \frac{\phi(|d|)}{|d|} \sum_{m_1 \le \sqrt{N}} \frac{1}{m_1} \left(\frac{d}{m_1}\right) + O(\sqrt{N}),$$

for fixed d and arbitrarily large N. Since $(d|m_1)$ is a nonprincipal character to the modulus $|d|$, the sum of its values as m_1 varies over any range is bounded. Hence the second double sum is $O(\sqrt{N})$. Thus

$$w^{-1} \sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n) = N \frac{\phi(|d|)}{|d|} \sum_{m \le \sqrt{N}} \frac{1}{m} \left(\frac{d}{m}\right) + O(\sqrt{N}).$$

We can extend the sum over N to infinity, and the remainder is estimated by

$$\sum_{m > \sqrt{N}} \frac{1}{m} \left(\frac{d}{m}\right) = O(N^{-\frac{1}{2}}),$$

on using partial summation. This again contributes an error $O(\sqrt{N})$ in the above asymptotic expression. In particular, we conclude that

(13) $$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n) = w \frac{\phi(|d|)}{|d|} \sum_{m=1}^{\infty} \frac{1}{m} \left(\frac{d}{m}\right).$$

Since $\phi(|d|)/|d|$ measures the density of the integers n for which $(n, d) = 1$, we can express the result in the form: The average with respect to n of $R(n)$ is $wL(1, \chi)$, where $\chi(m) = (d|m)$.

The next step is to evaluate the average of $R(n)$ from its original definition. Let $R(n,f)$ denote the number of representations of n (primary if $d > 0$) by a particular form f of discriminant d. Then

(14) $$R(n) = \sum_{f} R(n,f),$$

where the summation is over a representative set of forms (with $a > 0$), so that the number of terms in the sum is $h(d)$. We shall now evaluate

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n,f),$$

and it will turn out to be independent of f. Comparison of the two limits will give the relation between $h(d)$ and $L(1, \chi)$.

Take first the case $d < 0$. Then

$$\sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n,f)$$

is the number of pairs of integers x, y satisfying

$$0 < ax^2 + bxy + cy^2 \leq N, \qquad (ax^2 + bxy + cy^2, d) = 1.$$

The second condition limits x, y to certain pairs of residue classes to the modulus $|d|$, and it is easily proved[6] that the number of these pairs is $|d|\phi(|d|)$. Hence it suffices to consider the number of pairs of integers x, y satisfying

$$ax^2 + bxy + cy^2 \leq N, \qquad x \equiv x_0, \quad y \equiv y_0 \pmod{|d|}.$$

The first inequality expresses that the point (x, y) is in an ellipse with center at the origin, and as $N \to \infty$ this ellipse expands uniformly.

[6] Landau, *Vorlesungen*, Satz 206.

The area of the ellipse is

$$\frac{2\pi}{\sqrt{4ac - b^2}}N = \frac{2\pi}{|d|^{\frac{1}{2}}}N.$$

Intuition suggests—and a rigorous proof is easily given by dividing the plane into squares of side $|d|$—that the number of points is asymptotic to

$$\frac{1}{|d|^2}\frac{2\pi}{|d|^{\frac{1}{2}}}N$$

as $N \to \infty$. We have to multiply this by $|d|\phi(|d|)$ to allow for the various possibilities for x_0, y_0. Thus the conclusion is that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\substack{n=1\\(n,d)=1}}^{N} R(n,f) = \frac{\phi(|d|)}{|d|}\frac{2\pi}{|d|^{\frac{1}{2}}}.$$

Comparison with (13) and (14) gives

(15) $$h(d) = \frac{w|d|^{\frac{1}{2}}}{2\pi}L(1,\chi) \qquad \text{for } d < 0.$$

Now take the case $d > 0$. Arguing as before, we need the number of integer points (x, y) satisfying

$$ax^2 + bxy + cy^2 \leq N, \qquad x - \theta y > 0, \qquad 1 \leq \frac{x - \theta' y}{x - \theta y} < \varepsilon^2,$$

and

$$x \equiv x_0, \qquad y \equiv y_0 \qquad (\text{mod } d).$$

The first set of conditions represents a sector of a hyperbola bounded by two fixed lines (or rather half-lines) through the origin. The area of this sector is easily calculated by changing the coordinates from x, y to ξ, η, where

$$\xi = x - \theta y, \qquad \eta = x - \theta' y.$$

We have

$$\frac{\partial(\xi, \eta)}{\partial(x, y)} = \theta - \theta' = \frac{\sqrt{d}}{a}.$$

In the ξ, η plane, the sector is given by

$$\xi\eta \leq N/a, \qquad \xi > 0, \qquad \xi \leq \eta < \varepsilon^2\xi.$$

These conditions are equivalent to

$$0 < \xi \leq (N/a)^{\frac{1}{2}}, \qquad \xi \leq \eta < \min(\varepsilon^2\xi, N/a\xi).$$

Hence the area is

$$\int_0^{\xi_1} (\varepsilon^2\xi - \xi)\, d\xi + \int_{\xi_1}^{(N/a)^{1/2}} \left(\frac{N}{a\xi} - \xi\right) d\xi,$$

where $\xi_1 = \varepsilon^{-1}(N/a)^{\frac{1}{2}}$. This is

$$(\varepsilon^2 - 1)\tfrac{1}{2}\xi_1^2 + (N/a)^{\frac{1}{2}} \log(N/a) - (N/a)\log \xi_1 - \tfrac{1}{2}(N/a) + \tfrac{1}{2}\xi_1^2,$$

which reduces to

$$(N/a)\log \varepsilon.$$

This has to be divided by $d^{\frac{1}{2}}a^{-1}$ to give the area in the x, y plane. We have then to divide this by d^2 to allow for the congruences to the modulus d, and to multiply by $d\phi(d)$ to allow for the choices of x_0, y_0. This gives

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n=1 \\ (n,d)=1}}^{N} R(n,f) = \frac{\phi(d)}{d}\cdot\frac{\log\varepsilon}{d^{\frac{1}{2}}},$$

and comparison with (13) and (14) gives

$$h(d) = \frac{d^{\frac{1}{2}}}{\log\varepsilon} L(1,\chi) \qquad \text{for } d > 0. \tag{16}$$

This completes the first stage of the work, and, as we said earlier, the results (15) and (16) render visible the fact that $L(1,\chi) > 0$.

There remains the question of expressing $L(1,\chi)$ by means of a finite sum, as was done in §1 in the particular case when $|d|$ is a prime. The work is on the same general lines as there, but one needs the evaluation of a slight extension of Gauss' sum. This takes the form[7]

$$\sum_{m=1}^{|d|} \left(\frac{d}{m}\right) e(mn/|d|) = \left(\frac{d}{n}\right)\varepsilon|d|^{\frac{1}{2}},$$

where $\varepsilon = 1$ if $d > 0$ and $\varepsilon = i$ if $d < 0$. I will merely quote the final results[8]:

$$L(1,\chi) = -\frac{\pi}{|d|^{\frac{3}{2}}} \sum_{m=1}^{|d|} m\left(\frac{d}{m}\right) \qquad \text{if } d < 0, \tag{17}$$

$$L(1,\chi) = -\frac{1}{d^{\frac{1}{2}}} \sum_{m=1}^{d} \left(\frac{d}{m}\right) \log\sin\frac{m\pi}{d} \qquad \text{if } d > 0. \tag{18}$$

[7] Landau, *Vorlesungen*, Satz 215.
[8] Landau, *Vorlesungen*, Satz 217.

Here (17) is the more general form of (7) of §1 [there is similarly a more general form of (8)], and (18) is the more general form of (9) of §1.

The case when $d = -q$, where q is a prime congruent to 3 (mod 4), is particularly simple and interesting. We suppose that $q > 3$ so as to avoid any complication with the value of w. We have $w = 2$, and on combining (15) with (17) we get

$$h(d) = -\frac{1}{|d|}\sum_{m=1}^{q-1} m\left(\frac{-q}{m}\right) = -\frac{1}{q}\sum_{m=1}^{q-1} m\left(\frac{m}{q}\right). \tag{19}$$

It was this particular case of the class-number formula that was conjectured originally by Jacobi, and the considerations that led him to make the conjecture are curious. The number on the right of (19) is certainly an integer, say H, since by Euler's criterion

$$\sum_{m=1}^{q-1} m\left(\frac{m}{q}\right) \equiv \sum_{m=1}^{q-1} m^{\frac{1}{2}q+\frac{1}{2}} \equiv 0 \pmod{q}.$$

Jacobi proved, by an ingenious argument involving products and quotients of Gaussian sums, that H has the following property: for every prime $p \equiv 1 \pmod{q}$, there is a representation of $p^{|H|}$ in the form

$$4p^{|H|} = x^2 + qy^2.$$

(The reader will not be surprised to learn that Jacobi was unable to prove that H is positive.) On the other hand, it can be deduced from the theory of quadratic forms that the same property is possessed by the class number $h(-q)$. This led Jacobi to look for a connection between them, and after examining a number of particular cases he formulated the conjecture that $h(-q) = |H|$.

We conclude this section by stating briefly the connection between the theory of classes of equivalent quadratic forms and the theory of ideals in quadratic fields, but we shall omit the proofs.[9]

Let K be a quadratic field of discriminant d and let $\mathfrak{a}$ be an integral ideal in K. The general integer ξ of $\mathfrak{a}$ is given by

$$\xi = \alpha x + \beta y,$$

where α, β is a basis of $\mathfrak{a}$ and x, y run through all the rational integers. Thus

$$N\xi = (\alpha x + \beta y)(\alpha' x + \beta' y),$$

[9] See Landau, *Vorlesungen* III, pp. 186–198 or Hecke, *Algebraische Zahlen*, §53.

and this is a quadratic form in x and y with rational integral coefficients. All three coefficients are divisible by $N\mathfrak{a}$, and if we write

$$\frac{N\xi}{N\mathfrak{a}} = ax^2 + bxy + cy^2,$$

the discriminant of this form is d. The class to which the form belongs is independent of the choice of basis for the ideal, and is also the same for two equivalent ideals, provided that equivalence of ideals is defined in the narrow sense. That means that two ideals $\mathfrak{a}, \mathfrak{b}$ in K are said to be equivalent if there is a number λ of K with $N\lambda > 0$ such that

$$\mathfrak{a} = (\lambda)\mathfrak{b}$$

(of course λ need not be integral). Further, there is a one-to-one correspondence between a representative set of forms of discriminant d (positive definite if $d < 0$) and a representative set of ideals relative to equivalence in the narrow sense.

If $d < 0$ there is no distinction between equivalence in the narrow sense and in the ordinary sense, for then $N\lambda$ is necessarily positive. If $d > 0$ and there is a unit in K of norm -1, there is also no distinction, for we can ensure that $N\lambda > 0$ by multiplying λ by such a unit if necessary. If $d > 0$ but there is no unit of norm -1, each ideal class in the ordinary sense comprises two ideal classes in the narrow sense. It follows that, if we denote by $h_1(d)$ the number of ideal classes in K in the ordinary sense, then:

$$h(d) = h_1(d)$$

if $d < 0$ or if $d > 0$ and there is a unit in K of norm -1; but

$$h(d) = 2h_1(d)$$

if $d > 0$ and there is no unit of norm -1.

There is a similar one-to-one correspondence between the automorphs of the fields, when $d > 0$, and the units in K of norm $+1$. If ε_1 denotes the fundamental unit of K, then $\varepsilon = \varepsilon_1$ if $N\varepsilon_1 = +1$, but $\varepsilon = \varepsilon_1^2$ if $N\varepsilon_1 = -1$.

Combining these results, we have in both cases

$$h(d)\log\varepsilon = 2h_1(d)\log\varepsilon_1 \qquad \text{for } d > 0.$$

Thus the final expressions for the class number $h_1(d)$ of a quadratic field become

$$h_1(d) = -\frac{w}{2|d|}\sum_{m=1}^{|d|} m\left(\frac{d}{m}\right) \qquad \text{if } d < 0,$$

$$h_1(d)\log\varepsilon_1 = -\frac{1}{2}\sum_{m=1}^{d}\left(\frac{d}{m}\right)\log\sin\frac{m\pi}{d} \qquad \text{if } d > 0.$$

Dirichlet's class-number formula, as given in (15) and (16), can be regarded as a special case of a theorem[10] that applies to any algebraic number field K, by which the product of the class number and the regulator is expressed in terms of the residue at $s = 1$ of the Dedekind ζ function of K. If K is a quadratic field, the Dedekind ζ function is simply $\zeta(s)L(s, \chi)$, and the residue is $L(1, \chi)$. But this special case is of interest in its own right, particularly in view of the fact that $L(1, \chi)$ can be expressed by a finite sum, as in (17) and (18).

[10] Hecke, *Algebraische Zahlen*, Satz 125.

7

THE DISTRIBUTION OF THE PRIMES

Legendre was the first, as far as we know, to make any significant conjecture about the distribution of the primes. Let $\pi(x)$ denote the number of primes not exceeding x. Then Legendre conjectured, somewhat tentatively, that for large x the number $\pi(x)$ is given approximately by

$$\frac{x}{\log x - 1.08...}.$$

This would presumably imply, at the very least, that the ratio of $\pi(x)$ to $x/\log x$ tends to 1 as $x \to \infty$; and this is the celebrated Prime Number Theorem, which was first proved by Hadamard and de la Vallée Poussin independently in 1896. If we construe the conjecture in the more precise form that

$$\pi(x) = \frac{x}{\log x - A(x)},$$

where $A(x) \to 1.08...$ as $x \to \infty$, then it is erroneous, since (as we shall see) the limit of $A(x)$ is 1.

Gauss, in a letter of 1849 (which, however, was not published until much later), related that as a boy he had thought much on this question, and had reached the conclusion that a good approximation to $\pi(x)$ was given by

$$\operatorname{li} x = \int_2^x \frac{dt}{\log t}.$$

He certainly believed that the ratio $\pi(x)/\operatorname{li} x$ has the limit 1, which again is equivalent to the prime number theorem; how much more he believed is uncertain.[1] The asymptotic expansion of $\operatorname{li} x$, found by integrating by parts several times, is

$$\operatorname{li} x = \frac{x}{\log x} + \frac{1!x}{(\log x)^2} + \cdots + \frac{q!x}{(\log x)^{q+1}}[1 + \varepsilon(x)]$$

[1] See Landau, *Handbuch*, Kap. 1.

for any fixed q, where $\varepsilon(x) \to 0$ as $x \to \infty$. If the second term of this is significant in the approximation to $\pi(x)$ by li x, as we now know that it is, the limit of Legendre's $A(x)$ is 1.

The first mathematician of all time to prove any worthwhile results about the behavior of $\pi(x)$ as $x \to \infty$ was Tchebychev, in 1851 and 1852. In his first paper he provided some measure of justification for Gauss' conjectural association of $\pi(x)$ with li x. He proved that

$$\underline{\lim}\,\frac{\pi(x)}{\operatorname{li} x} \le 1 \le \overline{\lim}\,\frac{\pi(x)}{\operatorname{li} x},$$

so that if the limit exists it must be 1. But he further proved (in effect) that if there is a function with an asymptotic expansion of the same general character as li x which gives a good approximation to $\pi(x)$, then this function can only be li x itself. The proof is based on the asymptotic behavior of various combinations of $\zeta(s)$, $\zeta'(s)$, $\zeta''(s)$,... as $s \to 1$ from the right.[2]

In his second paper Tchebychev gave definite inequalities for $\pi(x)$; he proved that

$$(1) \qquad (0.92...)\frac{x}{\log x} < \pi(x) < (1.105...)\frac{x}{\log x}$$

for all sufficiently large x.

The proof depends on an interesting identity satisfied by the arithmetical function $\Lambda(n)$, which is defined by

$$(2) \qquad \Lambda(n) = \begin{cases} \log p & \text{if } n \text{ is a power of a prime } p, \\ 0 & \text{otherwise.} \end{cases}$$

The identity states that

$$(3) \qquad \sum_{m|n} \Lambda(m) = \log n.$$

Although this can be proved directly, the simplest way of deriving this and similar identities is by comparing coefficients in two Dirichlet series which have the same sum. By logarithmic differentiation of Euler's identity,

$$(4) \qquad -\frac{\zeta'(s)}{\zeta(s)} = \sum_p \sum_{m=1}^{\infty} (\log p)p^{-ms} = \sum_{n=1}^{\infty} \Lambda(n)n^{-s}.$$

[2] See Landau, *Handbuch*, Kap. 10.

Multiplying this by $\zeta(s)$, we get

$$\left(\sum \Lambda(n)n^{-s}\right)\left(\sum n^{-s}\right) = -\frac{\zeta'(s)}{\zeta(s)}\zeta(s) = \sum (\log n)n^{-s},$$

for $s > 1$, and comparison of coefficients gives (3).

If we sum (3) over positive integers $n \leq x$, we obtain

$$T(x) = \sum_{m \leq x} \Lambda(m)\left[\frac{x}{m}\right] = \sum_{n \leq x} \log n = \log[x]!,$$

and the number on the right is $x \log x - x + O(\log x)$ by Stirling's formula. This was the basis for Tchebychev's proof of (1).

A result of the same general character as (1), but with less precise constants, can be proved by considering the combination

$$T(x) - 2T(\tfrac{1}{2}x) = \sum_{m \leq x} \Lambda(m)\left(\left[\frac{x}{m}\right] - 2\left[\frac{x}{2m}\right]\right).$$

The left side is asymptotic to $x \log 2$, and the right side is

$$\leq \sum_{m \leq x} \Lambda(m) = \sum_{p \leq x} (\log p)\left[\frac{\log x}{\log p}\right] \leq (\log x)\pi(x).$$

This yields a lower bound for $\pi(x)$ of the desired character. The right side above is also

$$\geq \sum_{\frac{1}{2}x < m \leq x} \Lambda(m) \geq \sum_{\frac{1}{2}x < p \leq x} \log p \geq (\log \tfrac{1}{2}x)[\pi(x) - \pi(\tfrac{1}{2}x)].$$

This gives an upper bound for $\pi(x) - \pi(\frac{1}{2}x)$, from which an upper bound for $\pi(x)$ is easily derived by an inductive argument. Tchebychev's proof of (1) was based on the consideration of the more elaborate combination[3]

$$T(x) - T(\tfrac{1}{2}x) - T(\tfrac{1}{3}x) - T(\tfrac{1}{5}x) + T(\tfrac{1}{30}x).$$

The next substantial progress was made by Mertens in 1874. He proved that

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + A + O[(\log x)^{-1}], \tag{5}$$

a result which (even in a less precise form) had been attempted by Tchebychev without success. The proof, as one now sees it, is not particularly difficult. We have seen that

$$\sum_{m \leq x} \Lambda(m)\left[\frac{x}{m}\right] = T(x) = x \log x + O(x).$$

[3] Landau, *Handbuch*, Kap. 5.

The contribution of the prime values of m is

$$\sum_{p \leq x} (\log p)\left[\frac{x}{p}\right] = x \sum_{p \leq x} \frac{\log p}{p} + O[(\log x)\pi(x)]$$
$$= x \sum_{p \leq x} \frac{\log p}{p} + O(x).$$

Other values of m contribute $O(x)$. Hence

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1).$$

Denoting the sum on the left by $s(x)$, we have

$$\sum_{p \leq x} \frac{1}{p} = \sum_{2 \leq n \leq x} [s(n) - s(n-1)] \frac{1}{\log n},$$

and on applying partial summation we obtain (5).

Another result of Mertens is of interest in connection with Dirichlet's work on primes in an arithmetic progression. If χ is any nonprincipal character (mod q), it follows from the results of §4 that

$$\sum_p \frac{\chi(p)}{p^s}$$

has a finite limit as $s \to 1$ from the right; for the amount by which this series differs from $\log L(s, \chi)$ is trivial. Mertens proved the deeper result, which is suggested by the preceding one but cannot be deduced directly from it, that

$$\sum_p \frac{\chi(p)}{p} \tag{6}$$

converges. From this and (5) he easily deduced, by taking a linear combination of characters in the usual way, that

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod{q}}} \frac{1}{p} = \frac{1}{\phi(q)} \log\log x + A(q, a) + O[(\log x)^{-1}].$$

We have here a more precise form of Dirichlet's theorem that the series on the left, when extended to infinity, is divergent.

The proof of the convergence of the series (6) is simple and ingenious. Using (3), we have

$$\sum_{n \leq x} \frac{\chi(n) \log n}{n} = \sum_{m_1 m_2 \leq x} \frac{\chi(m_1)\chi(m_2)\Lambda(m_1)}{m_1 m_2}$$
$$= \sum_{m_1 \leq x} \frac{\chi(m_1)\Lambda(m_1)}{m_1} \sum_{m_2 \leq x/m_1} \frac{\chi(m_2)}{m_2}.$$

The inner sum on the right differs from $L(1, \chi)$ by a remainder that is $O(m_1/x)$, by partial summation. Hence the last expression is

$$L(1, \chi) \sum_{m \leq x} \frac{\chi(m)\Lambda(m)}{m} + O\left(x^{-1} \sum_{m \leq x} \Lambda(m)\right).$$

The last error term is $O[x^{-1}(\log x)\pi(x)] = O(1)$, by (1). Since the series $\Sigma\chi(n)(\log n)/n$ is convergent, by Dirichlet's test, it follows that

$$\sum_{m \leq x} \frac{\chi(m)\Lambda(m)}{m} = O(1),$$

and from this the convergence of the series (6) is deduced by partial summation.

It may be of interest to observe that the convergence of the series (6) implies the convergence of the Euler product for $L(s, \chi)$ when $s = 1$. Hence

$$L(1, \chi) = \prod_p [1 - \chi(p)p^{-1}]^{-1}.$$

8

RIEMANN'S MEMOIR

In his epoch-making memoir of 1860 (his only paper on the theory of numbers) Riemann showed that the key to the deeper investigation of the distribution of the primes lies in the study of $\zeta(s)$ as a function of the complex variable s. More than 30 years were to elapse, however, before any of Riemann's conjectures were proved, or any specific results about primes were established on the lines which he had indicated.

Riemann proved two main results:

(a) The function $\zeta(s)$ can be continued analytically over the whole plane and is then meromorphic, its only pole being a simple pole at $s = 1$ with residue 1. In other words, $\zeta(s) - (s - 1)^{-1}$ is an integral function.

(b) $\zeta(s)$ satisfies the functional equation

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = \pi^{-\frac{1}{2}(1-s)}\Gamma[\tfrac{1}{2}(1 - s)]\zeta(1 - s),$$

which can be expressed by saying that the function on the left is an even function of $s - \frac{1}{2}$. The functional equation allows the properties of $\zeta(s)$ for $\sigma < 0$ to be inferred from its properties for $\sigma > 1$. In particular, the only zeros of $\zeta(s)$ for $\sigma < 0$ are at the poles of $\Gamma(\frac{1}{2}s)$, that is, at the points $s = -2, -4, -6, \ldots$. These are called the *trivial zeros*. The remainder of the plane, where $0 \leq \sigma \leq 1$, is called *the critical strip*.

Riemann further made a number of remarkable conjectures.

(a′) $\zeta(s)$ has infinitely many zeros in the critical strip. These will necessarily be placed symmetrically with respect to the real axis, and also with respect to the central line $\sigma = \frac{1}{2}$ (the latter because of the functional equation).

(b′) The number $N(T)$ of zeros of $\zeta(s)$ in the critical strip with $0 < t \leq T$ satisfies the asymptotic relation

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T). \tag{1}$$

This was proved by von Mangoldt, first in 1895 with a slightly less good error term and then fully in 1905. We shall come to the proof in §15.

(c′) The integral function $\xi(s)$ defined by

$$\xi(s) = \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s)$$

(integral because it has no pole for $\sigma \geq \frac{1}{2}$ and is an even function of $s - \frac{1}{2}$) has the product representation

$$\xi(s) = e^{A+Bs}\prod_{\rho}\left(1 - \frac{s}{\rho}\right)e^{s/\rho}, \tag{2}$$

where A and B are constants and ρ runs through the zeros of $\zeta(s)$ in the critical strip. This was proved by Hadamard in 1893, as also was (a′) above. It played an important part in the proofs of the prime number theorem by Hadamard and de la Vallée Poussin. We shall come to the proof in §§ 11 and 12.

(d′) There is an explicit formula for $\pi(x) - \operatorname{li} x$, valid for $x > 1$, the most important part of which consists of a sum over the complex zeros ρ of $\zeta(s)$. As this is somewhat complicated to state, we give instead the closely related but somewhat simpler formula for $\psi(x) - x$, where

$$\psi(x) = \sum_{n \leq x} \Lambda(n). \tag{3}$$

It is:

$$\psi(x) - x = -\sum_{\rho}\frac{x^{\rho}}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \tfrac{1}{2}\log(1 - x^{-2}). \tag{4}$$

This was proved by von Mangoldt in 1895 (as was Riemann's original formula), and we give the proof in §17. In interpreting (4) two conventions have to be observed: first, in the sum over ρ the terms ρ and $\bar{\rho}$ are to be taken together, and second, if x is an integer, the last term $\Lambda(x)$ in the sum (3) defining $\psi(x)$ is to be replaced by $\frac{1}{2}\Lambda(x)$.

(e′) The famous Riemann Hypothesis, still undecided: that the zeros of $\zeta(s)$ in the critical strip all lie on the central line $\sigma = \frac{1}{2}$. It was proved by Hardy in 1914 that infinitely many of the zeros lie on the line, and by A. Selberg in 1942 that a positive proportion at least of all the zeros lie on the line.

There is very little indication of how Riemann was led to some of these conjectures. In 1932 Siegel[1] published an asymptotic expan-

[1] *Quellen und Studien zur Geschichte der Mathematik*, **2,** 45–80 (1932).

sion for $\zeta(s)$, valid in the critical strip, which had its origin in notes of Riemann preserved in the Göttingen University Library. From Siegel's description of the notes, it is plain that Riemann had more knowledge about $\zeta(s)$ than is apparent from his published memoir; but there is no reason to think that he had proofs of any of his conjectures.

In the present section we shall prove what Riemann proved, that is (in effect) the functional equation, and we shall follow one of his two methods. Many other proofs have since been given,[2] but this one is still the most elegant.

Riemann started from the classical definition of the Γ function:

$$\Gamma(\tfrac{1}{2}s) = \int_0^\infty e^{-t} t^{\frac{1}{2}s-1}\, dt,$$

valid for $\sigma > 0$. Putting $t = n^2\pi x$, we get

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)n^{-s} = \int_0^\infty x^{\frac{1}{2}s-1}e^{-n^2\pi x}\, dx.$$

Hence, for $\sigma > 1$,

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = \int_0^\infty x^{\frac{1}{2}s-1}\left(\sum_1^\infty e^{-n^2\pi x}\right) dx,$$

the inversion of order being justified by the convergence of

$$\sum_1^\infty \int_0^\infty x^{\frac{1}{2}\sigma-1}e^{-n^2\pi x}\, dx.$$

Writing

$$\omega(x) = \sum_1^\infty e^{-n^2\pi x},$$

we have

$$\begin{aligned}\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) &= \int_0^\infty x^{\frac{1}{2}s-1}\omega(x)\, dx \\ &= \int_1^\infty x^{\frac{1}{2}s-1}\omega(x)\, dx + \int_1^\infty x^{-\frac{1}{2}s-1}\omega(1/x)\, dx.\end{aligned}$$

Plainly

$$2\omega(x) = \theta(x) - 1,$$

[2] See Titchmarsh, Chap. 2.

where

$$\theta(x) = \sum_{-\infty}^{\infty} e^{-n^2\pi x}. \tag{5}$$

This function satisfies the simple functional equation

$$\theta(x^{-1}) = x^{\frac{1}{2}}\theta(x) \qquad \text{for } x > 0, \tag{6}$$

as we shall prove below; this equation is a special case of those satisfied by the ϑ functions of Jacobi. It follows that

$$\omega(x^{-1}) = -\tfrac{1}{2} + \tfrac{1}{2}x^{\frac{1}{2}} + x^{\frac{1}{2}}\omega(x).$$

Hence

$$\int_1^\infty x^{-\frac{1}{2}s-1}\omega(x^{-1})\,dx = \int_1^\infty x^{-\frac{1}{2}s-1}[-\tfrac{1}{2} + \tfrac{1}{2}x^{\frac{1}{2}} + x^{\frac{1}{2}}\,\omega(x)]\,dx$$
$$= -\frac{1}{s} + \frac{1}{s-1} + \int_1^\infty x^{-\frac{1}{2}s-\frac{1}{2}}\omega(x)\,dx.$$

We have therefore proved that

$$\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) = \frac{1}{s(s-1)} + \int_1^\infty (x^{\frac{1}{2}s-1} + x^{-\frac{1}{2}s-\frac{1}{2}})\omega(x)\,dx. \tag{7}$$

This holds for $\sigma > 1$. But the integral on the right converges absolutely for any s, and converges uniformly with respect to s in any bounded part of the plane, since

$$\omega(x) = O(e^{-\pi x})$$

as $x \to +\infty$. Hence the integral represents an everywhere regular function of s, and the above formula gives the analytic continuation of $\zeta(s)$ over the whole plane. It also gives the functional equation, since the right side is unchanged when s is replaced by $1 - s$.

We note that the function

$$\xi(s) = \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s)$$

is regular everywhere. Since $\frac{1}{2}s\Gamma(\frac{1}{2}s)$ has no zeros, the only possible pole of $\zeta(s)$ is at $s = 1$, and we have already seen (p. 32) that this is in fact a simple pole with residue 1.

Since $\Gamma(\frac{1}{2}s) \sim (\frac{1}{2}s)^{-1}$ as $s \to 0$, we deduce from (7) that $\zeta(0) = -\frac{1}{2}$. It is easily verified that

$$\omega(x) = e^{-\pi x} + e^{-4\pi x} + e^{-9\pi x} + \cdots < \tfrac{1}{2}x^{-\frac{1}{2}} \qquad \text{for } x > 1,$$

so if $0 < s < 1$ the integral in (7) is less than $\{s(1 - s)\}^{-1}$. Hence $\zeta(s) < 0$ for $0 < s < 1$. [The same conclusion may be drawn, more simply, from (7) of §4.]

It remains to prove the functional equation (6) of the θ function. We shall prove this in the more general form

$$(8) \qquad \sum_{-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} = x^{\frac{1}{2}} \sum_{-\infty}^{\infty} e^{-n^2\pi x + 2\pi i n\alpha},$$

which reduces to (6) when $\alpha = 0$, since we shall need this in the next section. It is supposed in (8) that $x > 0$ and that α is any real number (though actually the equation holds for complex x and α, provided $\Re x > 0$, with the value of $x^{-\frac{1}{2}}$ which has argument between $-\frac{1}{4}\pi$ and $\frac{1}{4}\pi$).

By Poisson's summation formula (§2),

$$\sum_{n=-N}^{N}{}' e^{-(n+\alpha)^2\pi/x} = \sum_{\nu=-\infty}^{\infty} \int_{-N}^{N} e^{-(t+\alpha)^2\pi/x + 2\pi i\nu t}\, dt.$$

Here we can replace N by ∞, since

$$\int_N^{\infty} e^{-(t+\alpha)^2\pi/x} \cos 2\pi\nu t\, dt = -\frac{1}{2\pi\nu}\int_N^{\infty} \sin 2\pi\nu t\, d[e^{-(t+\alpha)^2\pi/x}]$$

by integration by parts, and therefore

$$\left|\sum_{\nu\neq 0} \int_N^{\infty} e^{-(t+\alpha)^2\pi/x} \cos 2\pi\nu t\, dt\right| < Ce^{-(N+\alpha)^2\pi/x},$$

where C is a constant. Since this disappears as $N \to \infty$, the limit operation is justified. Thus

$$\begin{aligned}\sum_{-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} &= \sum_{\nu=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(t+\alpha)^2\pi/x + 2\pi i\nu t}\, dt \\ &= x \sum_{\nu=-\infty}^{\infty} e^{-2\pi i\nu\alpha} \int_{-\infty}^{\infty} e^{-\pi x u^2 + 2\pi i\nu x u}\, du.\end{aligned}$$

The quadratic in the exponent is

$$-\pi x(u - i\nu)^2 - \pi x\nu^2.$$

Now

$$\int_{-\infty}^{\infty} e^{-\pi x(u+\beta)^2}\, du = \int_{-\infty}^{\infty} e^{-\pi x v^2}\, dv = Ax^{-\frac{1}{2}},$$

where A is a positive constant; this holds for any β (real or complex) and simply expresses a movement in the path of integration from the real axis to another line parallel to it. Hence

$$\sum_{-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} = Ax^{\frac{1}{2}} \sum_{\nu=-\infty}^{\infty} e^{-\pi x\nu^2 - 2\pi i\nu\alpha}.$$

If we now take $\alpha = 0$ and apply this formula twice, we get $A^2 = 1$, whence $A = 1$. This proves (8), on replacing ν by $-\nu$ on the right.

9

THE FUNCTIONAL EQUATION OF THE L FUNCTIONS

The functional equation for Dirichlet's L functions was first given by Hurwitz in 1882 (*Werke* I, pp. 72–88), though he confined himself to real characters since he was primarily interested in L functions in relation to quadratic forms. He first obtained the functional equation for the more general ζ function $\zeta(s, w)$, which will be given below, and deduced that of the L functions from it. We shall follow the method used by de la Vallée Poussin in 1896, which is an extension of that of Riemann used in the preceding section.

The functional equation is valid only for primitive characters. We need the expression for $\chi(n)$ as a linear combination of imaginary exponentials $e_q(mn)$, which we used earlier in §1 [(4) and (5)] in the case when the character is the Legendre symbol.

For any character $\chi(n)$ to the modulus q, the Gaussian sum $\tau(\chi)$ is defined by

$$\tau(\chi) = \sum_{m=1}^{q} \chi(m)e_q(m). \tag{1}$$

If $(n, q) = 1$, then

$$\chi(n)\tau(\bar{\chi}) = \sum_{m=1}^{q} \bar{\chi}(m)\chi(n)e_q(m)$$

$$= \sum_{h=1}^{q} \bar{\chi}(h)e_q(nh), \tag{2}$$

on putting $m \equiv nh \pmod q$. This gives the desired expression for $\chi(n)$, provided that $(n, q) = 1$ and that $\tau(\chi) \neq 0$.

We now prove that, if χ is a primitive character, the last relation holds also when $(n, q) > 1$. We put

$$\frac{n}{q} = \frac{n_1}{q_1},$$

where $(n_1, q_1) = 1$ and $q_1 | q, q_1 < q$. We can suppose that $q_1 > 1$, since the relation holds trivially if n is a multiple of q. We have to prove that

$$\sum_{h=1}^{q} \bar{\chi}(h)e(n_1h/q_1) = 0.$$

Write $q = q_1q_2$ and put $h = uq_1 + v$, where

$$0 \leq u < q_2, \qquad 1 \leq v \leq q_1.$$

Then the exponential depends only on v, and it will suffice to prove that

$$\sum_{u=0}^{q_2-1} \bar{\chi}(uq_1 + v) = 0$$

for every v. Considered as a function of v, the last sum, say $S(v)$, is periodic with period q_1, for the effect of replacing v by $v + q_1$ is to change the range for u into $1 \leq u \leq q_2$, and $u = q_2$ is equivalent to $u = 0$. If c is any number satisfying

$$(c, q) = 1, \qquad c \equiv 1 \pmod{q_1}, \tag{3}$$

then

$$\chi(c)S(v) = \sum_{u=0}^{q_2-1} \bar{\chi}(cuq_1 + cv) = \sum_{u=0}^{q_2-1} \bar{\chi}(uq_1 + cv) = S(v). \tag{4}$$

We now appeal to the characteristic property of primitive characters (§5), namely that for $(n, q) = 1$, the function $\chi(n)$ is not periodic to any modulus q_1 that is a proper factor of q. This implies that there exist integers c_1, c_2 such that

$$(c_1, q) = (c_2, q) = 1, \qquad c_1 \equiv c_2 \pmod{q_1}, \qquad \chi(c_1) \neq \chi(c_2).$$

Hence there exists $c \equiv c_1c_2^{-1}$ which satisfies (3) and has $\chi(c) \neq 1$. It follows from (4) that $S(v) = 0$ for any v, as was to be proved.

We have proved that (2) holds independently of whether $(n, q) = 1$ or not. We now prove that, for a primitive character χ,

$$|\tau(\chi)| = q^{\frac{1}{2}}. \tag{5}$$

The proof given in §3 for a cubic character to a prime modulus applies equally to any nonprincipal (and therefore primitive) charac-

ter to a prime modulus, but does not readily extend to a composite modulus. The simplest proof is an indirect one. By (2),

$$|\chi(n)|^2|\tau(\chi)|^2 = \sum_{h_1=1}^{q} \sum_{h_2=1}^{q} \bar{\chi}(h_1)\chi(h_2)e_q[n(h_1 - h_2)].$$

Now sum for n over a complete set of residues (mod q). The sum of the values of $|\chi(n)|^2$ is $\phi(q)$, and the sum of the exponentials is 0 unless $h_1 \equiv h_2$. Hence

$$\phi(q)|\tau(\chi)|^2 = q \sum_{h} |\chi(h)|^2 = q\phi(q),$$

giving (5).

Although it is not necessary for our purpose, it may be of interest to evaluate $\tau(\chi)$ for a nonprimitive χ in terms of $\tau(\chi_1)$, where χ_1 is the primitive character (mod q_1) that induces χ. We have

$$\tau(\chi) = \sum_{m=1}^{q} \chi(m)e(m/q) = \sum_{\substack{m=1\\(m,q)=1}}^{q} \chi_1(m)e(m/q).$$

Put $q = q_1 r$. We first prove that $\tau(\chi) = 0$ if q_1 and r are not relatively prime. Put $D = (q_1, r)$; then the values of m that occur in the sum can be expressed as

$$m = m_1 + tq_1r/D,$$

where

$$(m_1, q) = 1, \qquad 0 < m_1 \le q_1r/D, \qquad 0 < t \le D.$$

But then $\chi_1(m) = \chi_1(m_1)$, since q_1r/D is an integral multiple of q_1. Hence the sum for $\tau(\chi)$ contains, as a factor, the sum

$$\sum_{t=1}^{D} e(t/D),$$

and this is 0 since $D > 1$. Thus it remains only to consider the case in which $(q_1, r) = 1$. Here we can put

$$m = uq_1 + vr, \qquad \text{where} \quad 0 < u \le r, \qquad 0 < v \le q_1.$$

This gives

$$\tau(\chi) = \sum_{\substack{u=1\\(u,r)=1}}^{r} \sum_{\substack{v=1\\(v,q_1)=1}}^{q_1} \chi_1(vr)e\left(\frac{u}{r} + \frac{v}{q_1}\right)$$

$$= \mu(r)\chi_1(r)\tau(\chi_1).$$

We can now rewrite (2) in the form

$$\chi(n) = \frac{1}{\tau(\bar{\chi})} \sum_{m=1}^{q} \bar{\chi}(m) e(mn/q). \tag{6}$$

The functional equation of an L function takes different forms according as $\chi(-1) = 1$ or $\chi(-1) = -1$. One of these must hold, since $\chi(-1)^2 = \chi(1) = 1$.

Suppose that $\chi(-1) = 1$. We have

$$\pi^{-\frac{1}{2}s} q^{\frac{1}{2}s} \Gamma(\tfrac{1}{2}s) n^{-s} = \int_0^\infty e^{-n^2\pi x/q} x^{\frac{1}{2}s-1}\, dx,$$

and on multiplying by $\chi(n)$ and summing over n we get

$$\pi^{-\frac{1}{2}s} q^{\frac{1}{2}s} \Gamma(\tfrac{1}{2}s) L(s, \chi) = \int_0^\infty x^{\frac{1}{2}s-1} \left[\sum_{n=1}^{\infty} \chi(n) e^{-n^2\pi x/q} \right] dx, \tag{7}$$

for $\sigma > 1$. Since $\chi(-1) = 1$ and $\chi(0) = 0$, we can write this as

$$\frac{1}{2} \int_0^\infty x^{\frac{1}{2}s-1} \psi(x, \chi)\, dx,$$

where

$$\psi(x, \chi) = \sum_{-\infty}^{\infty} \chi(n) e^{-n^2\pi x/q}.$$

A functional equation that relates $\psi(x, \chi)$ to $\psi(x^{-1}, \bar{\chi})$ can be deduced from (6) and the functional equation (8) of §8, with x replaced by x/q. We have

$$\begin{aligned}
\tau(\bar{\chi})\psi(x, \chi) &= \sum_{m=1}^{q} \bar{\chi}(m) \sum_{n=-\infty}^{\infty} e^{-n^2\pi x/q + 2\pi i mn/q} \\
&= \sum_{m=1}^{q} \bar{\chi}(m) (q/x)^{\frac{1}{2}} \sum_{n=-\infty}^{\infty} e^{-(n+m/q)^2 \pi q/x} \\
&= (q/x)^{\frac{1}{2}} \sum_{m=1}^{q} \bar{\chi}(m) \sum_{n=-\infty}^{\infty} e^{-(nq+m)^2\pi/xq} \\
&= (q/x)^{\frac{1}{2}} \sum_{l=-\infty}^{\infty} \bar{\chi}(l) e^{-l^2\pi/xq} \\
&= (q/x)^{\frac{1}{2}} \psi(x^{-1}, \bar{\chi}).
\end{aligned}$$

Now we split the integral in (7) into two parts, as in §8, and obtain

$$\pi^{-\frac{1}{2}s}q^{\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)L(s,\chi)$$

$$= \frac{1}{2}\int_1^\infty x^{\frac{1}{2}s-1}\psi(x,\chi)\,dx + \frac{1}{2}\int_1^\infty x^{-\frac{1}{2}s-1}\psi(x^{-1},\chi)\,dx$$

$$= \frac{1}{2}\int_1^\infty x^{\frac{1}{2}s-1}\psi(x,\chi)\,dx + \frac{1}{2}\frac{q^{\frac{1}{2}}}{\tau(\bar{\chi})}\int_1^\infty x^{-\frac{1}{2}s-\frac{1}{2}}\psi(x,\bar{\chi})\,dx.$$

This expression represents an everywhere regular function of s, and therefore gives the analytic continuation of $L(s,\chi)$ over the whole plane, regular everywhere since $\Gamma(\frac{1}{2}s)$ is never 0. Moreover, if we replace s by $1 - s$ and χ by $\bar{\chi}$, the above expression becomes

$$\frac{1}{2}\frac{q^{\frac{1}{2}}}{\tau(\chi)}\int_1^\infty x^{\frac{1}{2}s-1}\psi(x,\chi)\,dx + \frac{1}{2}\int_1^\infty x^{-\frac{1}{2}s-\frac{1}{2}}\psi(x,\bar{\chi})\,dx,$$

which is equal to the previous expression multiplied by $q^{\frac{1}{2}}/\tau(\chi)$, since

$$\tau(\chi)\tau(\bar{\chi}) = q.$$

The last relation is a consequence of (5) and $\chi(-1) = 1$, since the latter implies that $\overline{\tau(\chi)} = \tau(\bar{\chi})$.

We have now obtained the functional equation for $L(s,\chi)$ in the form

$$(8)\qquad \begin{cases} \pi^{-\frac{1}{2}(1-s)}q^{\frac{1}{2}(1-s)}\Gamma[\frac{1}{2}(1-s)]L(1-s,\bar{\chi}) \\ \quad = \dfrac{q^{\frac{1}{2}}}{\tau(\chi)}\pi^{-\frac{1}{2}s}q^{\frac{1}{2}s}\Gamma(\frac{1}{2}s)L(s,\chi), \end{cases}$$

and this is valid for any primitive character χ to the modulus q for which $\chi(-1) = 1$. Since $L(1-s,\bar{\chi})$ has no zeros for $1 - \sigma > 1$, that is, for $\sigma < 0$, and $\Gamma[\frac{1}{2}(1-s)]$ has no zeros at all, the only zeros of $L(s,\chi)$ for $\sigma < 0$ are at $s = -2, -4, -6, \ldots$, corresponding to the poles of $\Gamma(\frac{1}{2}s)$. There is also a zero of $L(s,\chi)$ at $s = 0$, corresponding to the pole of $\Gamma(\frac{1}{2}s)$ there.

Suppose that $\chi(-1) = -1$. The previous argument fails, since now the function $\psi(x,\chi)$ simply vanishes. We modify the procedure

by writing $\frac{1}{2}(s+1)$ in place of $\frac{1}{2}s$ in the original formula, so that this becomes

$$\pi^{-\frac{1}{2}(s+1)}q^{\frac{1}{2}(s+1)}\Gamma[\tfrac{1}{2}(s+1)]n^{-s} = \int_0^\infty ne^{-n^2\pi x/q}x^{\frac{1}{2}s-\frac{1}{2}}\,dx,$$

and gives

$$\pi^{-\frac{1}{2}(s+1)}q^{\frac{1}{2}(s+1)}\Gamma[\tfrac{1}{2}(s+1)]L(s,\chi) = \frac{1}{2}\int_0^\infty \psi_1(x,\chi)x^{\frac{1}{2}s-\frac{1}{2}}\,dx,$$

where

$$\psi_1(x,\chi) = \sum_{-\infty}^{\infty} n\chi(n)e^{-n^2\pi x/q}.$$

The functional equation satisfied by $\psi_1(x,\chi)$, analogous to that satisfied by $\psi(x,\chi)$, is

$$\tau(\bar{\chi})\psi_1(x,\chi) = iq^{\frac{1}{2}}x^{-\frac{3}{2}}\psi_1(x^{-1},\bar{\chi}), \tag{9}$$

and this is proved by the same reasoning as before, but with an appeal to the relation

$$\sum_{-\infty}^{\infty} ne^{-n^2\pi x/q+2\pi imn/q} = i(q/x)^{\frac{3}{2}}\sum_{-\infty}^{\infty}(n+m/q)e^{-\pi(n+m/q)^2q/x}. \tag{10}$$

The latter is deduced from (8) of §8 as follows. We have

$$\sum_{-\infty}^{\infty} e^{-n^2\pi y+2\pi in\alpha} = y^{-\frac{1}{2}}\sum_{-\infty}^{\infty} e^{-(n+\alpha)^2\pi/y}.$$

Differentiation with respect to α, justified by the uniform convergence of the differentiated series, gives

$$2\pi i\sum_{-\infty}^{\infty} ne^{-n^2\pi y+2\pi in\alpha} = -2\pi y^{-\frac{3}{2}}\sum_{-\infty}^{\infty}(n+\alpha)e^{-(n+\alpha)^2\pi/y},$$

and, on replacing y by x/q and α by m/q, we get (10).

Using (9) in the integral above, as in the preceding case, we obtain

$$\pi^{-\frac{1}{2}(s+1)}q^{\frac{1}{2}(s+1)}\Gamma[\tfrac{1}{2}(s+1)]L(s,\chi)$$
$$= \frac{1}{2}\int_1^\infty \psi_1(x,\chi)x^{\frac{1}{2}s-\frac{1}{2}}\,dx + \frac{1}{2}\frac{iq^{\frac{1}{2}}}{\tau(\bar{\chi})}\int_1^\infty \psi_1(x,\bar{\chi})x^{-\frac{1}{2}s}\,dx.$$

This again gives the continuation of $L(s,\chi)$ as a regular function over the whole plane. If we replace s by $1-s$ and χ by $\bar{\chi}$, the expression

becomes equal to its previous value multiplied by $iq^{\frac{1}{2}}/\tau(\chi)$, since now

$$\tau(\chi)\tau(\bar{\chi}) = -q.$$

Thus the functional equation in the present case takes the form

$$\pi^{-\frac{1}{2}(2-s)}q^{\frac{1}{2}(2-s)}\Gamma[\tfrac{1}{2}(2-s)]L(1-s,\bar{\chi}) = \frac{iq^{\frac{1}{2}}}{\tau(\chi)}\pi^{-\frac{1}{2}(s+1)}q^{\frac{1}{2}(s+1)}\Gamma[\tfrac{1}{2}(s+1)]L(s,\chi). \tag{11}$$

The zeros of $L(s,\chi)$ for $\sigma < 0$ are now at the poles of $\Gamma[\frac{1}{2}(s+1)]$, that is, at $s = -1, -3, -5, \ldots$.

It is possible to put together the two forms of the functional equation in (8) and (11) by introducing a number $\mathfrak{a}$, depending on χ, defined by

$$\mathfrak{a} = \begin{cases} 0 & \text{if } \chi(-1) = 1, \\ 1 & \text{if } \chi(-1) = -1. \end{cases} \tag{12}$$

Then the functional equation takes the form: if

$$\xi(s,\chi) = (\pi/q)^{-\frac{1}{2}(s+\mathfrak{a})}\Gamma[\tfrac{1}{2}(s+\mathfrak{a})]L(s,\chi), \tag{13}$$

then

$$\xi(1-s,\bar{\chi}) = \frac{i^{\mathfrak{a}}q^{\frac{1}{2}}}{\tau(\chi)}\xi(s,\chi). \tag{14}$$

Another method of proof, as mentioned at the beginning of this section, is to relate $L(s,\chi)$ to the function $\zeta(s,\alpha)$, which is defined for $0 < \alpha < 1$ by

$$\zeta(s,\alpha) = \sum_{n=0}^{\infty}(n+\alpha)^{-s}. \tag{15}$$

This reduces to $\zeta(s)$ when $\alpha = 1$ and to $(2^s - 1)\zeta(s)$ when $\alpha = \frac{1}{2}$. The relationship follows at once from the periodicity of $\chi(n)$: We have

$$L(s,\chi) = \sum_{m=1}^{q}\chi(m)\sum_{n=0}^{\infty}(qn+m)^{-s} = q^{-s}\sum_{m=1}^{q}\chi(m)\zeta(s, m/q). \tag{16}$$

The function $\zeta(s,\alpha)$, like $\zeta(s)$, can be continued to be regular everywhere except for a simple pole at $s = 1$. For $\sigma < 0$, it is expressible

in terms of two other convergent Dirichlet series, with $1 - s$ in place of s, by the formula[1]

$$\zeta(s, \alpha) = \frac{2\Gamma(1 - s)}{(2\pi)^{1-s}}\left(\sin \tfrac{1}{2}\pi s \sum_{1}^{\infty} \frac{\cos 2\pi m\alpha}{m^{1-s}} + \cos \tfrac{1}{2}\pi s \sum_{1}^{\infty} \frac{\sin 2\pi m\alpha}{m^{1-s}}\right).$$

The use of this relation in (16) leads again to the functional equation for $L(s,\chi)$, though in an unsymmetric form.

There is nothing corresponding to a Euler product for $\zeta(s, \alpha)$, except when $\alpha = 1$ or $\frac{1}{2}$, and it behaves in many ways quite differently from $\zeta(s)$. Heilbronn and I proved[2] that, if α is rational ($\neq 1$ or $\frac{1}{2}$) or transcendental, then $\zeta(s, \alpha)$ has zeros in $\sigma > 1$, and Cassels[3] proved the same in the more difficult case when α is an algebraic irrational.

[1] See, for example, Titchmarsh, §2.17.
[2] *J. London Math. Soc.*, **11**, 181–185 (1936).
[3] *J. London Math. Soc.*, **36**, 177–184 (1961).

10

PROPERTIES OF THE Γ FUNCTION

We collect some properties of the Γ function for convenience of reference.[1] The usual definition is by means of Euler's integral:

$$(1) \qquad \Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, dt,$$

but this applies only for $\sigma > 0$. Weierstrass' formula

$$(2) \qquad \frac{1}{s\Gamma(s)} = e^{\gamma s} \prod_{n=1}^{\infty} (1 + s/n) e^{-s/n},$$

where γ is Euler's constant, applies in the whole plane, and shows that $\Gamma(s)$ has no zeros and has simple poles at $s = 0, -1, -2, \ldots$.

Among the functional relations satisfied by $\Gamma(s)$ are

$$(3) \qquad \begin{gathered} \Gamma(s+1) = s\Gamma(s), \\ \Gamma(s)\Gamma(1-s) = \pi/\sin \pi s, \qquad \Gamma(s)\Gamma(s+\tfrac{1}{2}) = 2^{1-2s}\pi^{\frac{1}{2}}\Gamma(2s), \end{gathered}$$

the last being Legendre's duplication formula. Combined, they give

$$\Gamma(\tfrac{1}{2}s)/\Gamma(\tfrac{1}{2} - \tfrac{1}{2}s) = \pi^{-\frac{1}{2}} 2^{1-s} \cos \tfrac{1}{2}s\pi\Gamma(s),$$

and if this is used in the functional equation of $\zeta(s)$ (p. 59), it gives the unsymmetric form of the functional equation:

$$(4) \qquad \zeta(1-s) = 2^{1-s}\pi^{-s}(\cos \tfrac{1}{2}s\pi)\Gamma(s)\zeta(s).$$

Stirling's asymptotic formula, in the simple form

$$(5) \qquad \log \Gamma(s) = (s - \tfrac{1}{2})\log s - s + \tfrac{1}{2}\log 2\pi + O(|s|^{-1}),$$

is valid as $|s| \to \infty$, in the angle $-\pi + \delta < \arg s < \pi - \delta$, for any fixed $\delta > 0$. Under the same conditions,

$$(6) \qquad \frac{\Gamma'(s)}{\Gamma(s)} = \log s + O(|s|^{-1}).$$

[1] Proofs will be found in many books, e.g., in Whittaker and Watson, *Modern Analysis*, Chaps. 12 and 13. See also Ingham, footnote on p. 57, with reference to (6).

11

INTEGRAL FUNCTIONS OF ORDER 1

The next important progress in the theory of the ζ function, after Riemann's pioneering paper, was made by Hadamard, who developed the theory of integral functions of finite order in the early 1890's and applied it to $\zeta(s)$ via $\xi(s)$. His results were used in both the proofs of the prime number theorem, given by himself and by de la Vallée Poussin, though later it was found that for the particular purpose of proving the prime number theorem, they could be dispensed with.

An integral function $f(z)$ is said to be of finite order if there exists a number α such that

$$(1) \qquad f(z) = O(e^{|z|^\alpha}) \qquad \text{as } |z| \to \infty.$$

We must have $\alpha > 0$, excluding the case when $f(z)$ is just a constant. The lower bound of the numbers α with the property (1) is called the *order* of $f(z)$.

An integral function of finite order with no zeros is necessarily of the form $e^{g(z)}$, where $g(z)$ is a polynomial, and its order is simply the degree of $g(z)$ and so is an integer. For $g(z) = \log f(z)$ can be defined so as to be single valued, and is itself an integral function. It satisfies

$$\Re g(z) = \log|f(z)| < 2R^\alpha$$

on any large circle $|z| = R$. If we put

$$g(z) = \sum_0^\infty (a_n + ib_n)z^n,$$

then

$$\Re g(z) = \sum_0^\infty a_n R^n \cos n\theta - \sum_1^\infty b_n R^n \sin n\theta,$$

for $z = Re^{i\theta}$. If we assume $g(0) = 0$, as we may, then

$$\pi|a_n|R^n \le \int_0^{2\pi} |\Re g(Re^{i\theta})| d\theta = \int_0^{2\pi} \{|\Re g(Re^{i\theta})| + \Re g(Re^{i\theta})\} d\theta < 8\pi R^\alpha.$$

It follows, on making $R \to \infty$, that $a_n = 0$ if $n > \alpha$, and similarly for b_n. This proves that $g(z)$ is a polynomial, and it is then obvious that the order of $f(z)$ is equal to the degree of $g(z)$.

We observe, for future reference, that in the preceding argument it suffices if the estimate for $f(z)$ on $|z| = R$ holds for some sequence of values of R with limit infinity, instead of for all large R.

Now suppose that an integral function $f(z)$ of finite order ρ has zeros at $z_1, z_2, \ldots$ (multiple zeros being repeated as appropriate). The question arises: How is the distribution of the zeros related to the order ρ? This question is most easily answered by means of Jensen's formula[1]: if $z_1, \ldots, z_n$ are the zeros of $f(z)$ in $|z| < R$, and there is no zero on $|z| = R$, then

$$(2) \qquad \frac{1}{2\pi}\int_0^{2\pi} \log|f(Re^{i\theta})|\, d\theta - \log|f(0)| = \log\frac{R^n}{|z_1|\ldots|z_n|}.$$

[We suppose, for convenience, that $f(0) \neq 0$.] An alternative expression for the right side is

$$\int_0^R r^{-1}n(r)\, dr,$$

where $n(r)$ denotes the number of zeros in $|z| < r$. For if $|z_1| = r_1$, and so on, the value of the integral is

$$\log r_2/r_1 + 2\log r_3/r_2 + \cdots + n\log R/r_n = \log(R^n/r_1 r_2 \ldots r_n).$$

Jensen's formula is easily established by factorizing $f(z)$ as

$$(z - z_1)\ldots(z - z_n)F(z)$$

and proving the formula for each factor separately.

It follows from Jensen's formula that the zeros of an integral function of given order ρ cannot be too dense. For if $\alpha > \rho$, we have

$$\log|f(Re^{i\theta})| < R^\alpha$$

for all sufficiently large R, whence

$$\int_0^R r^{-1}n(r)\, dr < R^\alpha - \log|f(0)| < 2R^\alpha.$$

Since

$$\int_R^{2R} r^{-1}n(r)\, dr \geq n(R)\int_R^{2R} r^{-1}\, dr = n(R)(\log 2),$$

[1] Strangely enough, Jensen's formula was not discovered until after the work of Hadamard.

it follows that

$$n(R) = O(R^\alpha). \tag{3}$$

A consequence of this estimate is that $\sum r_n^{-\beta}$ converges if $\beta > \alpha$, and therefore converges if $\beta > \rho$. For

$$\sum_1^\infty r_n^{-\beta} = \int_0^\infty r^{-\beta}\, dn(r) = \beta \int_0^\infty r^{-\beta-1} n(r)\, dr < \infty.$$

We are now in a position to represent $f(z)$ by a simple canonical product, of the kind introduced by Weierstrass. From now on we suppose that $\rho = 1$, since this is the only case with which we shall be concerned later. We can then assert that $\Sigma r_n^{-1-\varepsilon}$ converges for any $\varepsilon > 0$, and in particular that Σr_n^{-2} converges. Hence the product

$$P(z) = \prod_{n=1}^\infty (1 - z/z_n)e^{z/z_n}$$

(if it does not terminate) converges absolutely for all z, and converges uniformly in any bounded domain not containing any of the points z_n. Hence it represents an integral function with zeros (of the appropriate multiplicities) at $z_1, z_2, \ldots$. If we put

$$f(z) = P(z)F(z), \tag{4}$$

then $F(z)$ is an integral function without zeros.

We cannot immediately conclude that $F(z) = e^{g(z)}$, where $g(z)$ is a polynomial, because it is not obvious that $F(z)$ is of finite order. The most direct way of proving the desired result is to obtain a lower bound for $|P(z)|$, and hence an upper bound for $|F(z)|$, on a sequence of circles $|z| = R$, and then appeal to the result proved earlier. The values of R must be kept away from the numbers r_n. Since Σr_n^{-2} converges, the total length of all the intervals $(r_n - r_n^{-2}, r_n + r_n^{-2})$ on the real line is finite, and consequently there exist arbitrarily large values of R with the property that

$$|R - r_n| > r_n^{-2} \qquad \text{for all } n. \tag{5}$$

Put $P(z) = P_1(z)P_2(z)P_3(z)$, where these are the subproducts extended over the following sets of n:

$$P_1: \quad |z_n| < \tfrac{1}{2}R,$$
$$P_2: \quad \tfrac{1}{2}R \le |z_n| \le 2R,$$
$$P_3: \quad |z_n| > 2R.$$

For the factors of P_1 we have, on $|z| = R$,

$$|(1 - z/z_n)e^{z/z_n}| \geq (|z/z_n| - 1)e^{-|z|/|z_n|} > e^{-R/r_n},$$

and since

$$\sum_{r_n < \frac{1}{2}R} r_n^{-1} < (\tfrac{1}{2}R)^{\varepsilon} \sum_{n=1}^{\infty} r_n^{-1-\varepsilon},$$

it follows that

$$|P_1(z)| > \exp(-R^{1+2\varepsilon}).$$

For the factors of P_2, we have

$$|(1 - z/z_n)e^{z/z_n}| \geq e^{-2}|z - z_n|/2R > CR^{-3},$$

where C is a positive constant, by (5). The number of factors is less than $R^{1+\varepsilon}$, by (3). Hence

$$|P_2(z)| > (CR^{-3})^{R^{1+\varepsilon}} > \exp(-R^{1+2\varepsilon}).$$

Finally, for the factors of P_3, we have

$$|(1 - z/z_n)e^{z/z_n}| > e^{-c(R/r_n)^2}$$

for some positive constant c, since $|z/z_n| < \frac{1}{2}$. We also have

$$\sum_{r_n > 2R} r_n^{-2} < (2R)^{-1+\varepsilon} \sum_{n=1}^{\infty} r_n^{-1-\varepsilon},$$

and therefore

$$|P_3(z)| > \exp(-R^{1+2\varepsilon}).$$

It follows that, on $|z| = R$, we have

$$|P(z)| > \exp(-R^{1+3\varepsilon}),$$

whence

$$|F(z)| < \exp(R^{1+4\varepsilon})$$

by (1) and (4). By what was proved earlier, this inequality, since it holds for a sequence of values of R with limit infinity, implies that $F(z) = e^{g(z)}$, where $g(z)$ is a polynomial of degree at most 1. Finally we have, therefore,

$$f(z) = e^{A+Bz} \prod_{n=1}^{\infty} (1 - z/z_n)e^{z/z_n}, \tag{6}$$

where A and B are constants.

We know that the series $\Sigma r_n^{-1-\varepsilon}$ converges for any $\varepsilon > 0$. The series Σr_n^{-1} may or may not converge, but if it does then $f(z)$ satisfies the inequality

$$|f(z)| < e^{C|z|} \tag{7}$$

for some constant C. This follows at once from the inequality (valid for all ζ)

$$|(1 - \zeta)e^{\zeta}| \leq e^{2|\zeta|},$$

which itself follows from the power series for $(1 - \zeta)e^{\zeta}$.

To summarize the results of the present section:

An integral function of order 1 *necessarily has the form* (6). *If* $r_n = |z_n|$, *where the* z_n *are the zeros of* $f(z)$, *then* $\Sigma r_n^{-1-\varepsilon}$ *converges for any* $\varepsilon > 0$. *If* Σr_n^{-1} *converges, then* $f(z)$ *satisfies* (7).

12

THE INFINITE PRODUCTS FOR $\xi(s)$ AND $\xi(s,\chi)$

We apply the conclusions of the preceding section to the integral function

(1) $$\xi(s) = \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s).$$

We first prove that

(2) $$|\xi(s)| < \exp(C|s|\log|s|) \qquad \text{as } |s| \to \infty,$$

for some constant C; this will establish that $\xi(s)$ is of order 1 at most. Since $\xi(s) = \xi(1-s)$, it will suffice to prove the inequality when $\sigma \geq \frac{1}{2}$. Obviously[1]

$$|\tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}| < \exp(C|s|),$$

and

$$|\Gamma(\tfrac{1}{2}s)| < \exp(C|s|\log|s|)$$

by Stirling's formula, which is applicable since $-\frac{1}{2}\pi < \arg s < \frac{1}{2}\pi$. Thus it remains to estimate $\zeta(s)$, and this is possible on the basis of the representation obtained in (7) of §4, namely

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty (x-[x])x^{-s-1}\,dx,$$

valid for $\sigma > 0$. The integral is bounded for $\sigma \geq \frac{1}{2}$, and therefore

(3) $$|\zeta(s)| < C|s|$$

when $|s|$ is large. This completes the proof of (2).

We see further that, as $s \to +\infty$ through real values, the inequality (2) is substantially (that is, apart from the value of C) the best possible, since $\log\Gamma(s) \sim s\log s$ and $\zeta(s) \to 1$. Consequently $\xi(s)$ does *not* satisfy the more precise inequality (7) of the preceding section.

[1] The constant C is not necessarily the same at each occurrence.

It follows that $\xi(s)$ has an infinity of zeros, say $\rho_1, \rho_2, \ldots$, such that

(4) $$\sum |\rho_n|^{-1-\varepsilon} \quad \text{converges for any } \varepsilon > 0$$

and

(5) $$\sum |\rho_n|^{-1} \quad \text{diverges;}$$

and that

(6) $$\xi(s) = e^{A+Bs} \prod_\rho (1 - s/\rho)e^{s/\rho}.$$

The zeros of $\xi(s)$ are the nontrivial zeros of $\zeta(s)$, for in (1) the trivial zeros of $\zeta(s)$ are cancelled by the poles of $\Gamma(\frac{1}{2}s)$, and $\frac{1}{2}s\Gamma(\frac{1}{2}s)$ has no zeros, and the zero of $s - 1$ is cancelled by the pole of $\zeta(s)$. Hence $\zeta(s)$ *has an infinity of nontrivial zeros* ρ *in the critical strip* $0 \leq \sigma \leq 1$, *and these have the properties* (4) *and* (5).

The product formula (6) leads to an expression for $\zeta'(s)/\zeta(s)$ as a sum of partial fractions. Logarithmic differentiation of (6) gives

(7) $$\frac{\xi'(s)}{\xi(s)} = B + \sum_\rho \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right),$$

and, combined with the logarithmic derivative of (1), this gives

(8) $$\frac{\zeta'(s)}{\zeta(s)} = B - \frac{1}{s-1} + \frac{1}{2}\log \pi - \frac{1}{2}\frac{\Gamma'(\frac{1}{2}s+1)}{\Gamma(\frac{1}{2}s+1)} + \sum_\rho \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right).$$

This exhibits the pole of $\zeta(s)$ at $s = 1$ and the nontrivial zeros at $s = \rho$. The trivial zeros at $s = -2, -4, \ldots$ are contained in the Γ term, since

(9) $$-\frac{1}{2}\frac{\Gamma'(\frac{1}{2}s+1)}{\Gamma(\frac{1}{2}s+1)} = \frac{1}{2}\gamma + \sum_{n=1}^{\infty} \left(\frac{1}{s+2n} - \frac{1}{2n}\right)$$

by logarithmic differentiation from (2) of §10. The representation of ζ'/ζ in (8) will be the basis for much of the later work on $\zeta(s)$.

The constants A and B, though not very important, can be evaluated. By (1),

$$\xi(1) = \tfrac{1}{2}\pi^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}) \lim_{s\to 1} (s-1)\zeta(s) = \tfrac{1}{2},$$

whence $\xi(0) = \frac{1}{2}$ and therefore $e^A = \frac{1}{2}$ by (6).

As regards B, we have

$$B = \frac{\xi'(0)}{\xi(0)} = -\frac{\xi'(1)}{\xi(1)}$$

from (7) and the functional equation $\xi(s) = \xi(1 - s)$. By (1),

$$\frac{\xi'(s)}{\xi(s)} = \frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1} - \tfrac{1}{2}\log\pi + \frac{1}{2}\frac{\Gamma'(\tfrac{1}{2}s+1)}{\Gamma(\tfrac{1}{2}s+1)}.$$

It follows from (9) and the series for log 2 that

$$-\frac{1}{2}\frac{\Gamma'(\tfrac{3}{2})}{\Gamma(\tfrac{3}{2})} = \tfrac{1}{2}\gamma - 1 + \log 2.$$

Hence

$$B = \tfrac{1}{2}\gamma - 1 + \tfrac{1}{2}\log 4\pi - \lim_{s\to 1}\left[\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}\right].$$

To evaluate the limit, we have recourse again to (7) of §4:

$$\zeta(s) = \frac{s}{s-1} - sI(s), \qquad I(s) = \int_1^\infty (x - [x])x^{-s-1}\,dx.$$

A simple calculation shows that

$$\lim_{s\to 1}\left[\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}\right] = 1 - I(1).$$

Now

$$I(1) = \int_1^\infty (x - [x])x^{-2}\,dx = \lim_{N\to\infty}\left[\log N - \sum_{n=1}^{N-1} n\left(\frac{1}{n} - \frac{1}{n+1}\right)\right]$$

$$= \lim_{N\to\infty}\left(\log N - \sum_{n=1}^{N} n^{-1} + 1\right) = 1 - \gamma.$$

Hence

$$B = -\tfrac{1}{2}\gamma - 1 + \tfrac{1}{2}\log 4\pi. \tag{10}$$

We can give another interpretation for B, as follows. Although the series $\Sigma\,|\rho|^{-1}$ diverges, the series $\Sigma\rho^{-1}$ converges, provided one groups together the terms from ρ and $\bar{\rho}$. For if $\rho = \beta + i\gamma$, then

$$\frac{1}{\rho} + \frac{1}{\bar{\rho}} = \frac{2\beta}{\beta^2 + \gamma^2} \le \frac{2}{|\rho|^2},$$

and we know that $\Sigma|\rho|^{-2}$ converges. It follows from (7) and the functional equation for $\xi(s)$ that

$$B + \sum_{\rho}\left(\frac{1}{1-s-\rho} + \frac{1}{\rho}\right) = -B - \sum_{\rho}\left(\frac{1}{s-\rho} + \frac{1}{\rho}\right),$$

and the terms containing $1 - s - \rho$ and $s - \rho$ cancel, since if ρ is a zero then so is $1 - \rho$. Thus

$$(11) \qquad B = -\sum_{\rho}\frac{1}{\rho} = -2\sum_{\gamma>0}\frac{\beta}{\beta^2+\gamma^2}.$$

The numerical value of B is about -0.023; from this it can easily be seen that $|\gamma| > 6$ for all zeros.

We now apply similar considerations to the L functions. Let χ be a primitive character to the modulus q, and define, as in (13) of §9,

$$(12) \qquad \xi(s,\chi) = (q/\pi)^{\frac{1}{2}s+\frac{1}{2}\mathfrak{a}}\Gamma(\tfrac{1}{2}s + \tfrac{1}{2}\mathfrak{a})L(s,\chi),$$

where $\mathfrak{a}$ is 0 or 1 as in (12) of §9. [Note that there is no need to include the factor $s(s-1)$, which was inserted in the definition of $\xi(s)$ to cancel the poles of $\Gamma(\frac{1}{2}s)$ and $\zeta(s)$ at $s = 0$ and $s = 1$ respectively.] As we saw in §9, $\xi(s,\chi)$ is an integral function and satisfies the functional equation

$$(13) \qquad \xi(1-s,\bar{\chi}) = \frac{i^{\mathfrak{a}}q^{\frac{1}{2}}}{\tau(\chi)}\xi(s,\chi),$$

in which the multiplying factor has absolute value 1.

We need first an estimate for $L(s,\chi)$ when $|s|$ is large, and this is deduced on the same lines as for $\zeta(s)$, starting from (8) of §4. This states that

$$L(s,\chi) = s\int_1^{\infty} S(x)x^{-s-1}\,dx, \qquad \text{where } S(x) = \sum_{n\le x}\chi(n),$$

and is valid for $\sigma > 0$. Since $|S(x)| \le q$, it implies that

$$(14) \qquad |L(s,\chi)| \le 2q|s| \qquad \text{for } \sigma \ge \tfrac{1}{2}.$$

Hence

$$(15) \qquad \begin{aligned} |\xi(s,\chi)| &\le 2q^{\frac{1}{2}\sigma+\frac{1}{2}}|s|\,|\Gamma[\tfrac{1}{2}(s+\mathfrak{a})]| \\ &< q^{\frac{1}{2}\sigma+\frac{1}{2}}\exp(C|s|\log|s|) \end{aligned}$$

when $|s|$ is large. A similar result holds for $\sigma \le \frac{1}{2}$, by the functional equation. Again this inequality is substantially the best possible as $s \to +\infty$ through real values, since then $L(s,\chi) \to 1$. We conclude, as for $\zeta(s)$, that $L(s,\chi)$ *has an infinity of zeros* ρ *in the critical strip* $0 \le \sigma \le 1$, *which have the properties* (4) *and* (5). We also have

$$(16) \qquad \xi(s,\chi) = e^{A+Bs}\prod_{\rho}(1 - s/\rho)e^{s/\rho},$$

but A and B will now depend on χ. One can express $e^A = \xi(0,\chi)$ in terms of $\xi(1,\bar{\chi})$ and therefore in terms of $L(1,\bar{\chi})$.

The analog of (8), obtained by logarithmic differentiation from (16) and (12), is

$$\text{(17)}\quad \frac{L'(s,\chi)}{L(s,\chi)} = -\tfrac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})}{\Gamma(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})} + B(\chi) + \sum_{\rho}\left(\frac{1}{s-\rho}+\frac{1}{\rho}\right).$$

This, again, is the basis for much of the later work.

The number $B(\chi)$ can be expressed in terms of the expansion of L'/L in powers of s, but it seems to be very difficult to estimate $B(\chi)$ at all satisfactorily as a function of q. (In subsequent arguments it will usually be eliminated from the above equation by subtraction.) If we argue as in the proof of (11), we get

$$B(\chi) = \frac{\xi'(0,\chi)}{\xi(0,\chi)} = -\frac{\xi'(1,\bar{\chi})}{\xi(1,\bar{\chi})}$$

$$= -B(\bar{\chi}) - \sum_{\bar{\rho}}\left(\frac{1}{1-\bar{\rho}}+\frac{1}{\bar{\rho}}\right).$$

As $B(\bar{\chi}) = \overline{B(\chi)}$, it follows that

$$2\,\Re B(\chi) = -\sum_{\rho}\left(\Re\frac{1}{1-\bar{\rho}} + \Re\frac{1}{\bar{\rho}}\right).$$

We now write ρ in place of $1-\bar{\rho}$; this is permissible since permutation of non-negative terms does not alter a sum. Hence

$$\text{(18)}\qquad \Re B(\chi) = -\frac{1}{2}\sum_{\rho}\left(\frac{1}{\rho}+\frac{1}{\bar{\rho}}\right) = -\sum_{\rho}\Re\frac{1}{\rho}.$$

In particular, if χ is a real character, $B(\chi)$ is negative and is expressed in terms of the zeros ρ by (11). The difficulty of estimating $B(\chi)$ is connected with the fact that, as far as we know, $L(s,\chi)$ may have a zero near to $s=0$.

We observe that, for a complex χ, the zeros of $L(s,\chi)$ are still symmetric about the line $\sigma = \frac{1}{2}$, since $1-\bar{\rho} = \rho'$, but not about the real axis.

13

A ZERO-FREE REGION FOR $\zeta(s)$

It was proved independently by Hadamard and de la Vallée Poussin in 1896 that $\zeta(s) \neq 0$ on $\sigma = 1$. This was a vital step in their proofs of the prime number theorem, and it remained a vital step in all subsequent proofs until the discovery of an elementary proof[1] by Selberg and Erdös in 1948.

For $\sigma > 1$, we have

$$\log \zeta(s) = \sum_p \sum_{m=1}^{\infty} m^{-1} p^{-m\sigma} e^{-itm \log p}.$$

If $\zeta(s)$ had a zero at $1 + it$, then $\Re \log \zeta(\sigma + it)$ would tend to $-\infty$ as $\sigma \to 1$ from the right. This suggests that the numbers $\cos(tm \log p)$ would be predominantly negative. But then we should expect the numbers $\cos(2tm \log p)$ to be predominantly positive, and it seems likely that this would contradict the fact that $\Re \log \zeta(\sigma + 2it)$ remains bounded above as $\sigma \to 1$.

The line of reasoning just indicated was worked out in rigorous detail by Hadamard and (somewhat differently) by de la Vallée Poussin. Mertens[2] put the proof in a more elegant form by employing the inequality

(1) $$3 + 4\cos\theta + \cos 2\theta \geq 0,$$

which holds for all θ because the left side is $2(1 + \cos\theta)^2$. Applied to

$$\Re \log \zeta(s) = \sum_p \sum_{m=1}^{\infty} m^{-1} p^{-m\sigma} \cos(t \log p^m)$$

with t replaced by $0, t, 2t$ in succession, it gives

$$3 \log \zeta(\sigma) + 4\Re \log \zeta(\sigma + it) + \Re \log \zeta(\sigma + 2it) \geq 0.$$

Hence

(2) $$\zeta^3(\sigma)|\zeta^4(\sigma + it)\zeta(\sigma + 2it)| \geq 1$$

[1] For an account of this, see Hardy and Wright, *Introduction to the Theory of Numbers* (4th ed.), Chap. 22.

[2] *Sitzungsber. Akad. Wiss. Wien., Math.—Naturwiss. Classe*, **107**, 1429–1434 (1898).

for $\sigma > 1$. As $\sigma \to 1$, we have $\zeta(\sigma) \sim (\sigma - 1)^{-1}$. If $\zeta(1 + it) = 0$ for some t (which is necessarily not 0), then

$$|\zeta(\sigma + it)| < A(\sigma - 1)$$

for some constant A, as $\sigma \to 1$. Since $\zeta(\sigma + 2it)$ remains bounded as $\sigma \to 1$, we get a contradiction to the inequality (2). It will be seen that the success of the proof depends on the fact that the coefficient 4 in (1) is greater than the coefficient 3.

The argument was extended by de la Vallée Poussin in 1899 to show that $\zeta(s) \neq 0$ in a thin region to the left of $\sigma = 1$, the breadth of which at height t is proportional to $(\log t)^{-1}$ for large t. In proving this, it is more convenient to work with the function $\zeta'(s)/\zeta(s)$ than with the function $\log \zeta(s)$, since the analytic continuation of the latter to the left of $\sigma = 1$ is obviously difficult, whereas the former has its only poles for $\sigma > 0$ at the zeros of $\zeta(s)$. By logarithmic differentiation of the Euler product, as in (4) of §7, we have

$$-\Re\zeta'(s)/\zeta(s) = \sum_{n=1}^{\infty} \Lambda(n) n^{-\sigma} \cos(t \log n)$$

for $\sigma > 1$. Hence, by the same argument as before,

$$(3) \quad 3\left[-\frac{\zeta'(\sigma)}{\zeta(\sigma)}\right] + 4\left[-\Re\frac{\zeta'(\sigma + it)}{\zeta(\sigma + it)}\right] + \left[-\Re\frac{\zeta'(\sigma + 2it)}{\zeta(\sigma + 2it)}\right] \geq 0.$$

The behavior of $-\zeta'(\sigma)/\zeta(\sigma)$ as $\sigma \to 1$ from the right presents no difficulty; in view of the simple pole of $\zeta(s)$ at $s = 1$, we have

$$-\frac{\zeta'(\sigma)}{\zeta(\sigma)} < \frac{1}{\sigma - 1} + A$$

for $1 < \sigma \leq 2$, where A denotes a positive absolute constant (not necessarily the same at each occurrence).

The behavior of the other two functions near $\sigma = 1$ is obviously much influenced by any zero that $\zeta(s)$ may have just to the left of $\sigma = 1$, at a height near to t or $2t$. This influence is rendered explicit by the partial fraction formula

$$-\frac{\zeta'(s)}{\zeta(s)} = \frac{1}{s - 1} - B - \tfrac{1}{2}\log \pi + \frac{1}{2}\frac{\Gamma'(\frac{1}{2}s + 1)}{\Gamma(\frac{1}{2}s + 1)} - \sum_{\rho}\left(\frac{1}{s - \rho} + \frac{1}{\rho}\right),$$

which was (8) of §12. The Γ term is less than $A \log t$ if $t \geq 2$ and $1 \leq \sigma \leq 2$. Hence, in this region,

$$(4) \qquad -\Re\frac{\zeta'(s)}{\zeta(s)} < A \log t - \sum_{\rho} \Re\left(\frac{1}{s - \rho} + \frac{1}{\rho}\right).$$

The sum over ρ is positive, since

$$\Re\frac{1}{s-\rho}=\frac{\sigma-\beta}{|s-\rho|^2} \quad \text{and} \quad \Re\frac{1}{\rho}=\frac{\beta}{|\rho|^2}.$$

We obtain a valid inequality when $s = \sigma + 2it$ by just omitting the sum:

$$-\Re\frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)} < A\log t. \tag{5}$$

As regards $s = \sigma + it$, we choose t to coincide with the ordinate γ of a zero $\beta + i\gamma$, with $\gamma \geq 2$, and take just the one term $1/(s - \rho)$ in the sum which corresponds to this zero:

$$-\Re\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} < A\log t - \frac{1}{\sigma-\beta}.$$

Substituting these upper bounds in the basic inequality (3), we obtain

$$\frac{4}{\sigma-\beta} < \frac{3}{\sigma-1} + A\log t.$$

Take $\sigma = 1 + \delta/\log t$, where δ is a positive constant. Then

$$\beta < 1 + \frac{\delta}{\log t} - \frac{4\delta}{(3+A\delta)\log t},$$

and if δ is suitably chosen in relation to A, this gives

$$\beta < 1 - \frac{c}{\log t},$$

where c is a positive constant to which a numerical value could be assigned. Thus we have proved:

There exists a positive numerical constant c such that $\zeta(s)$ has no zero in the region

$$\sigma \geq 1 - \frac{c}{\log t}, \qquad t \geq 2.$$

In view of the fact that $\zeta(s)$ has no zero arbitrarily near $\sigma = 1$ with $|t| \leq 2$, we can also say that there exists a positive constant c such that $\zeta(s)$ has no zero in the region

$$\sigma \geq 1 - \frac{c}{\log(|t|+2)}.$$

The breadth of the zero-free region was enlarged to

$$\frac{c \log \log t}{\log t}$$

by Littlewood in 1922, and to[3]

$$\frac{c(\alpha)}{(\log t)^{\alpha}}$$

for any $\alpha > \frac{2}{3}$, by Vinogradov and Korobov independently in 1958. These improvements depend on upper bounds for $\zeta(s)$ in a region just to the left of $\sigma = 1$, which are deduced from somewhat elaborate estimations of exponential sums.[4]

[3] For the sake of simplicity, I give a slightly weakened version of the result.

[4] For an account, see A. Walfisz, *Weylsche Exponentialsummen in der neueren Zahlentheorie*, Berlin, 1963, Chaps. 2 and 5.

14

ZERO-FREE REGIONS FOR $L(s, \chi)$

There is no difficulty in extending the results of the preceding section to the zeros of $L(s, \chi)$ when χ is a *fixed* character. But this is of limited value; for many purposes it is important to allow q to vary and to have estimates that are explicit in respect of q. This raises some difficult problems, and the results so far known are better for complex characters than for real characters.

We no longer suppose that $t \geq 2$ but merely that $t \geq 0$. There is no loss of generality in the latter supposition, for the zeros of $L(s, \chi)$ with $t < 0$ are the complex conjugates of the zeros of $L(s, \bar{\chi})$ with $t > 0$. We are concerned with nonprincipal characters only, and therefore $q \geq 3$ throughout.

Logarithmic differentiation of the Euler product formula gives

$$-\frac{L'(s, \chi)}{L(s, \chi)} = \sum_{n=1}^{\infty} \Lambda(n) n^{-\sigma} \chi(n) e^{-it \log n}$$

for $\sigma > 1$. We can represent the real part of $\chi(n)e^{-it \log n}$, for $(n, q) = 1$, as $\cos \theta$, and θ has to be replaced by 2θ if χ is replaced by χ^2 and t by $2t$, and has to be replaced by 0 if χ is replaced by χ_0 and t by 0. Hence the analog of the inequality (3) of the preceding section is

$$(1) \qquad 3\left[-\frac{L'(\sigma, \chi_0)}{L(\sigma, \chi_0)}\right] + 4\left[-\Re \frac{L'(\sigma + it, \chi)}{L(\sigma + it, \chi)}\right] + \left[-\Re \frac{L'(\sigma + 2it, \chi^2)}{L(\sigma + 2it, \chi^2)}\right] \geq 0.$$

If χ is a real character (but only then) we have $\chi^2 = \chi_0$, and this affects the argument. The effect is important only when t is small and we come under the influence of the pole of $L(s, \chi_0)$ at $s = 1$.

We suppose first that χ is a complex primitive character, and follow as closely as possible the argument of the preceding section.

Again the first term presents no difficulty; we have

$$-\frac{L'(\sigma,\chi_0)}{L(\sigma,\chi_0)} = \sum_1^\infty \chi_0(n)\Lambda(n)n^{-\sigma} \le -\frac{\zeta'(\sigma)}{\zeta(\sigma)} < \frac{1}{\sigma-1} + c_1$$

for $1 < \sigma < 2$, where c_1 denotes a positive absolute constant (and similarly for $c_2, \ldots$ later[1]).

For the other two terms in (1), we have recourse to (17) of §12. This gives

$$-\Re\frac{L'(s,\chi)}{L(s,\chi)} = \tfrac{1}{2}\log\frac{q}{\pi} + \frac{1}{2}\Re\frac{\Gamma'(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})}{\Gamma(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})}$$
$$-\Re B(\chi) - \Re\sum_\rho\left(\frac{1}{s-\rho}+\frac{1}{\rho}\right),$$

where $\mathfrak{a}$ is 0 or 1. We can eliminate $B(\chi)$ and $\Sigma 1/\rho$ by appealing to (18) of §12. Since the Γ term above is $O[\log(t+2)]$, we can express the result in the form

$$(2)\qquad -\Re\frac{L'(s,\chi)}{L(s,\chi)} < c_2\mathscr{L} - \sum_\rho \Re\frac{1}{s-\rho},$$

where we have written for brevity

$$(3)\qquad \mathscr{L} = \log q + \log(t+2).$$

This holds (for $\sigma > 1$) for any primitive χ, whether real or complex. Since

$$\Re\frac{1}{s-\rho} = \frac{\sigma-\beta}{|s-\rho|^2} \ge 0,$$

we can as before omit the series or any part of it.

We omit the whole series when estimating $L'(\sigma+2it,\chi^2)/L(\sigma+2it,\chi^2)$. There is the minor complication that χ^2, though nonprincipal, may not be primitive. However, if χ_1 is the primitive character that induces χ^2, it follows from (3) of §5 that

$$\left|\frac{L'(s,\chi^2)}{L(s,\chi^2)} - \frac{L'(s,\chi_1)}{L(s,\chi_1)}\right| \le \sum_{p|q}\frac{p^{-\sigma}\log p}{1-p^{-\sigma}}$$
$$\le \sum_{p|q}\log p \le \log q.$$

Hence the upper bound in (2), namely $c_2\mathscr{L}$, remains valid.

[1] To leave the constants unnumbered, as we have done hitherto, would lead to confusion in the present section.

We choose t to be the ordinate γ of a zero $\beta + i\gamma$ of $L(s, \chi)$, and by retaining on the right of (2) only the one term corresponding to this zero, we obtain

$$-\Re \frac{L'(\sigma + it, \chi)}{L(\sigma + it, \chi)} < c_2\mathscr{L} - \frac{1}{\sigma - \beta}.$$

The three estimates, when substituted in the basic inequality (1), give

$$\frac{4}{\sigma - \beta} < \frac{3}{\sigma - 1} + c_3\mathscr{L}.$$

We take $\sigma = 1 + c_4/\mathscr{L}$, with a suitable c_4, and by the same argument as in the preceding section we obtain

$$\beta < 1 - c_5/\mathscr{L}. \tag{4}$$

This has been proved for any complex primitive χ, but the restriction to primitive χ can be removed, since, by (3) of §5, any zeros of $L(s, \chi)$ additional to those of $L(s, \chi_1)$, where χ_1 induces χ, are the zeros of a finite number of factors $1 - \chi_1(p)p^{-s}$ and are on $\sigma = 0$. We can accordingly assert that *there exists a positive absolute constant* c_5 *such that, if* χ *is a complex character to the modulus* q, *any zero* $\beta + i\gamma$ *of* $L(s, \chi)$ *satisfies* (4), *where*

$$\mathscr{L} = \log q + \log(|\gamma| + 2). \tag{5}$$

[We have modified the definition of $\mathscr{L}$ in (3) to accord with the choice $t = \gamma$.]

Suppose next that χ is a real primitive character. The preceding argument needs modification only in one respect: The inequality for $-\Re L'/L$ with $s = \sigma + 2it$ and χ replaced by χ^2 is no longer applicable since χ is the principal character. We must now relate L'/L to ζ'/ζ, and by the same argument as that used above when χ^2 was imprimitive, we have

$$\left|\frac{L'(s, \chi_0)}{L(s, \chi_0)} - \frac{\zeta'(s)}{\zeta(s)}\right| \le \log q$$

for $\sigma > 1$. As regards $-\zeta'/\zeta$, we cannot quote the inequality (5) of §13, because this was proved only for large t. In proving it, a term $1/(s - 1)$ was neglected. When this is restored, the same argument as was used there gives

$$-\Re \frac{\zeta'(s)}{\zeta(s)} < \Re \frac{1}{s - 1} + c_6 \log (t + 2).$$

Hence

$$-\Re\frac{L'(\sigma + 2it, \chi^2)}{L(\sigma + 2it, \chi^2)} < \Re\frac{1}{\sigma - 1 + 2it} + c_7\mathscr{L},$$

where $\mathscr{L}$ is again defined by (3).

Using this in place of the stronger inequality that was available for complex χ, we deduce from (1) that

$$\frac{4}{\sigma - \beta} < \frac{3}{\sigma - 1} + \Re\left(\frac{1}{\sigma - 1 + 2it}\right) + c_8\mathscr{L},$$

where now $t = \gamma$. If we take $\sigma = 1 + \delta/\mathscr{L}$, and postulate that $\gamma \geq \delta/\mathscr{L}$, we get

$$\frac{4}{\sigma - \beta} < \frac{3\mathscr{L}}{\delta} + \frac{\mathscr{L}}{5\delta} + c_8\mathscr{L},$$

whence

$$\beta < 1 - \frac{4 - 5c_8\delta}{16 + 5c_8\delta}\frac{\delta}{\mathscr{L}}.$$

If δ is sufficiently small in relation to c_8, we get an inequality of the form (4) but subject to the condition $\gamma \geq \delta/\mathscr{L}$, where $\mathscr{L}$ is given by (5). This condition is satisfied if $\gamma \geq \delta/\log q$. We have therefore proved that *there exists a positive absolute constant c_9 such that, if $0 < \delta < c_9$ and χ is a real nonprincipal character to the modulus q, then any zero $\beta + i\gamma$ of $L(s, \chi)$ for which*

$$|\gamma| \geq \frac{\delta}{\log q}$$

satisfies

$$\beta < 1 - \frac{\delta}{5\mathscr{L}},$$

where $\mathscr{L}$ is given by (5). We have omitted the requirement that χ should be primitive, for the same reason as before.

It remains to consider what can be proved about the zeros of $L(s, \chi)$, for real nonprincipal χ, with

$$|t| < \frac{\delta}{\log q},$$

where δ is a small positive constant. We shall show that there is at most one zero with $\sigma > 1 - \delta'/\log q$ for a suitable positive constant δ' and that, if there is one, it must be real. The final clause

is in fact a corollary, for if there were a nonreal zero, there would be two zeros at conjugate complex points.

The inequality (2), with $s = \sigma > 1$, can be written

$$-\frac{L'(\sigma, \chi)}{L(\sigma, \chi)} < c_{10} \log q - \sum_{\rho} \frac{1}{\sigma - \rho},$$

the last sum being real since the zeros occur in conjugate complex pairs. In quoting this inequality we have assumed, as we may without loss of generality, that χ is primitive. If there were zeros at $\beta \pm i\gamma$, where $\gamma \neq 0$, we should have

$$-\frac{L'(\sigma, \chi)}{L(\sigma, \chi)} < c_{10} \log q - \frac{2(\sigma - \beta)}{(\sigma - \beta)^2 + \gamma^2}.$$

For the left side, there is the crude lower bound

$$-\frac{L'(\sigma, \chi)}{L(\sigma, \chi)} = \sum_1^{\infty} \chi(n)\Lambda(n)n^{-\sigma} \geq -\sum_1^{\infty} \Lambda(n)n^{-\sigma} = \frac{\zeta'(\sigma)}{\zeta(\sigma)} > -\frac{1}{\sigma - 1} - c_{11}.$$

Thus

$$-\frac{1}{\sigma - 1} < c_{12} \log q - \frac{2(\sigma - \beta)}{(\sigma - \beta)^2 + \gamma^2}.$$

We take $\sigma = 1 + 2\delta/\log q$; then

$$|\gamma| < \frac{\delta}{\log q} = \tfrac{1}{2}(\sigma - 1) < \tfrac{1}{2}(\sigma - \beta),$$

and the last inequality implies that

$$-\frac{1}{\sigma - 1} < c_{12} \log q - \frac{8}{5(\sigma - \beta)}.$$

If the δ of the previous result is sufficiently small in relation to c_{12}, we get $\beta < 1 - \delta/\log q$.

The argument is substantially the same if, instead of two conjugate complex zeros, there are two real zeros (or a double real zero). Thus we have proved: *There exists a positive absolute constant c_{13} such that, if $0 < \delta < c_{13}$, the only possible zero of $L(s, \chi)$ for a real nonprincipal χ, satisfying*

$$|\gamma| < \frac{\delta}{\log q}, \qquad \beta > 1 - \frac{\delta}{\log q}$$

is a single (simple) real zero.

The three results proved so far can be fitted together to give the following theorem, which we state for convenience of reference. It simplifies the statement to consider two cases according as $|t| \geq 1$ or $|t| < 1$, since in the former case the number $\mathscr{L}$ is essentially $\log q|t|$ and in the latter case it is essentially $\log q$.

THEOREM. *There exists a positive absolute constant* c_{14} *with the following property. If* χ *is a complex character modulo q, then* $L(s, \chi)$ *has no zero in the region defined by*

$$\sigma \geq \begin{cases} 1 - \dfrac{c_{14}}{\log q|t|} & \text{if } |t| \geq 1, \\ 1 - \dfrac{c_{14}}{\log q} & \text{if } |t| \leq 1. \end{cases} \tag{6}$$

If χ *is a real nonprincipal character, the only possible zero of* $L(s, \chi)$ *in this region is a single* (*simple*) *real zero.*

These results are due partly to Gronwall[2] and partly to Titchmarsh.[3]

We shall now prove a result due to Landau,[4] the effect of which is to assure us that if there exist values of q for which an L function formed with a real primitive character $(\text{mod}\, q)$ has a zero with $\beta > 1 - c/\log q$, then such values of q are very rare. He proved that *if* χ_1, χ_2 *are distinct real primitive characters to the moduli* q_1, q_2 *respectively, and if the corresponding L functions have real zeros* β_1, β_2, *then*

$$\min(\beta_1, \beta_2) < 1 - \frac{c_{15}}{\log q_1 q_2}, \tag{7}$$

where c_{15} *is some positive absolute constant.* The possibility that $q_1 = q_2$ is not excluded.

The proof is based on the fact that $\chi_1(n)\chi_2(n)$ is a character to the modulus $q_1 q_2$, being multiplicative and periodic. It is not in general a primitive character, but it is nonprincipal. For if $\chi_1(n)\chi_2(n) = 1$ whenever $(n, q_1 q_2) = 1$, we should have $\chi_1(n) = \chi_2(n)$ whenever $(n, q_1 q_2) = 1$, and this would mean that the primitive characters χ_1 and χ_2 would induce the same character to the modulus $q_1 q_2$. This is impossible by the results of §5.

[2] *Rendiconti di Palermo*, **35**, 145–159 (1913).
[3] *Rendiconti di Palermo*, **54**, 414–429 (1930); **57**, 478–479 (1933).
[4] *Göttinger Nachrichten*, **1918**, 285–295.

For $\sigma > 1$, we have

$$-\frac{L'(\sigma, \chi_1\chi_2)}{L(\sigma, \chi_1\chi_2)} < c_{16} \log q_1 q_2;$$

this is proved by the same argument as that which we applied to $L(s, \chi^2)$ when χ^2 was nonprincipal but not necessarily primitive. Further, by (2),

$$-\frac{L'(\sigma, \chi_j)}{L(\sigma, \chi_j)} < c_{17} \log q_j - \frac{1}{\sigma - \beta_j},$$

the symbol $\mathfrak{R}$ in (2) being now superfluous.

Now consider the expression

$$-\frac{\zeta'(\sigma)}{\zeta(\sigma)} - \frac{L'(\sigma, \chi_1)}{L(\sigma, \chi_1)} - \frac{L'(\sigma, \chi_2)}{L(\sigma, \chi_2)} - \frac{L'(\sigma, \chi_1\chi_2)}{L(\sigma, \chi_1\chi_2)}$$

$$= \sum_{n=1}^{\infty} \Lambda(n)[1 + \chi_1(n)][1 + \chi_2(n)]n^{-\sigma} \geq 0.$$

On substituting the previous upper bounds, and also that for $-\zeta'(\sigma)/\zeta(\sigma)$, we get

$$\frac{1}{\sigma - \beta_1} + \frac{1}{\sigma - \beta_2} < \frac{1}{\sigma - 1} + c_{18} \log q_1 q_2.$$

If σ is taken to be $1 + \delta/\log q_1 q_2$, for a sufficiently small positive constant δ, the last inequality shows that β_1 and β_2 cannot both be greater than $1 - \delta'/\log q_1 q_2$, for a suitable positive δ'. This proves (7).

Various deductions can be made from the last result. In particular one sees that *for at most one of the real nonprincipal characters* χ *(mod* q*) can* $L(s, \chi)$ *have a zero in the region* (6). [We assume here tacitly that c_{14}, in the definition (6), is diminished if necessary so as to satisfy $c_{14} \leq \frac{1}{2}c_{15}$.]

Another deduction concerns the possible sequence $q_1, q_2, \ldots$ of positive integers q with the property that there is a real primitive χ (mod q) for which $L(s, \chi)$ has a real zero β satisfying

(8) $$\beta > 1 - c_{19}/\log q.$$

If c_{19} is suitably chosen, say $c_{19} = \frac{1}{3}c_{15}$, then

$$q_{j+1} > q_j^2.$$

For (7) implies that

$$1 - \frac{c_{19}}{\log q_j} < 1 - \frac{c_{15}}{\log q_j q_{j+1}},$$

whence the result.

A deduction made by Page[5] and applied by him to the prime number theorem for arithmetic progressions (see §20) concerns the set of positive integers $q \leq z$, where $z \geq 3$. *If c_{20} is a suitable positive constant, there is at most one real primitive χ to a modulus $q \leq z$ for which $L(s, \chi)$ has a real zero β satisfying*

$$\beta > 1 - \frac{c_{20}}{\log z}. \tag{9}$$

The last inequality is, of course, of a somewhat more stringent nature than (8). The proof is immediate; if there were two such characters, both the zeros would satisfy

$$\beta > 1 - \frac{c_{20}}{\log z} \geq 1 - \frac{2c_{20}}{\log q_1 q_2},$$

and this would contradict (7) if $c_{20} = \frac{1}{2}c_{15}$.

If there is such an "exceptional" real character χ_1 to a modulus $q_1 \leq z$, then q_1 will be a function of z, and the only real nonprincipal characters χ to moduli $q \leq z$ for which $L(s, \chi)$ has a real zero satisfying (9) will be χ_1 and the imprimitive characters induced by χ_1. Their moduli will be multiples of q_1.

The only obvious general upper bound for a real zero β of an L function corresponding to a real primitive χ is that which can be derived from the lower bound for $L(1, \chi)$ provided by the class-number formula. Since $h(d) \geq 1$, the formulas (15) and (16) of §6, in which $d = \pm q$, give

$$L(1, \chi) > c_{21} q^{-\frac{1}{2}}. \tag{10}$$

We can easily prove that

$$|L'(\sigma, \chi)| < c_{22} \log^2 q \qquad \text{for } 1 - 1/\log q \leq \sigma \leq 1, \tag{11}$$

and it then follows, by the mean value theorem of the differential calculus, that

$$L(1, \chi) = L(1, \chi) - L(\beta, \chi) < (1 - \beta) c_{22} \log^2 q,$$

[5] *Proc. London Math. Soc.*, (2)**39**, 116–141 (1935).

whence

$$\beta < 1 - \frac{c_{23}}{q^{\frac{1}{2}} \log^2 q}. \tag{12}$$

By the usual argument, this holds also for real nonprincipal χ, even if imprimitive. If $\chi(-1) = 1$, which corresponds to the case $d > 0$, the inequality (12) can be improved[6] to the extent of a factor $\log q$, since in (16) of §6 there is a factor $\log \varepsilon$, and $\log \varepsilon \geq \log \frac{1}{2}(1 + q^{\frac{1}{2}})$.

The proof of (11) is as follows. We have

$$L'(\sigma, \chi) = -\sum_{1}^{\infty} \chi(n)(\log n)n^{-\sigma}$$

for $\sigma > 0$. Since $n^{-\sigma} = e^{-\sigma \log n} \leq en^{-1}$ for $n \leq q$, we have

$$\left| \sum_{n=1}^{q} \chi(n)(\log n)n^{-\sigma} \right| \leq e \sum_{n=1}^{q} (\log n)n^{-1} < c_{24} \log^2 q.$$

By partial summation, noting that $(\log n)n^{-\sigma}$ decreases for $n > q$, we have

$$\begin{aligned} \left| \sum_{n=q+1}^{\infty} \chi(n)(\log n)n^{-\sigma} \right| &\leq (\log q)q^{-\sigma} \max_{N} \left| \sum_{q+1}^{N} \chi(n) \right|. \\ &\leq (\log q)eq^{-1}q. \end{aligned}$$

These results imply (11).

We remark, for convenience of reference, that the same argument applied to $L(\sigma, \chi)$ gives

$$|L(\sigma, \chi)| < c_{25} \log q \qquad \text{for } 1 - 1/\log q \leq \sigma \leq 1. \tag{13}$$

In §21 we shall prove a theorem due to Siegel, which establishes a much more precise inequality for a real zero than that given in (12). But whereas all the results of the present section have been "effective," in the sense that numerical values could be assigned to all the constants, it does not seem to be possible to derive an effective inequality from Siegel's theorem.

[6] Goldfeld and Schinzel (*Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, (4) **2**, 571–583 (1975)), have shown that $\beta < 1 - cq^{-1/2}$ if $\chi(-1) = -1$, and that $\beta < 1 - cq^{-1/2} \log g$ if $\chi(-1) = 1$.

15

THE NUMBER $N(T)$

In this section we prove the approximate formula for $N(T)$, the number of zeros of $\zeta(s)$ in the rectangle $0 < \sigma < 1$, $0 < t < T$, which was stated by Riemann and established by von Mangoldt. It was stated as (1) in §8.

It is convenient to work initially with $\xi(s)$ rather than with $\zeta(s)$ because of its simple functional equation. Assuming for simplicity that T (which we suppose to be large) does not coincide with the ordinate of a zero, we have

$$2\pi N(T) = \Delta_R \arg \xi(s),$$

where R is the rectangle in the s plane with vertices at

$$2, \qquad 2 + iT, \qquad -1 + iT, \qquad -1,$$

described in the positive sense. There is no change in $\arg \xi(s)$ as s describes the base of the rectangle, since $\xi(s)$ is then real and nowhere 0. Further, the change as s moves from $\frac{1}{2} + iT$ to $-1 + iT$ and then to -1 is equal to the change as s moves from 2 to $2 + iT$ and then to $\frac{1}{2} + iT$, since

$$\xi(\sigma + it) = \xi(1 - \sigma - it) = \overline{\xi(1 - \sigma + it)}.$$

Hence

$$\pi N(T) = \Delta_L \arg \xi(s),$$

where L denotes the line from 2 to $2 + iT$ and then to $\frac{1}{2} + iT$.

The definition of $\xi(s)$, in (1) of §12, can be written as

$$\xi(s) = (s - 1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s + 1)\zeta(s).$$

We have

$$\Delta_L \arg (s - 1) = \arg(iT - \tfrac{1}{2}) = \tfrac{1}{2}\pi + O(T^{-1}),$$
$$\Delta_L \arg \pi^{-\frac{1}{2}s} = \Delta_L(-\tfrac{1}{2}t \log \pi) = -\tfrac{1}{2}T \log \pi.$$

Also, by Stirling's formula (§10),

$$\begin{aligned}\Delta_L \arg\Gamma(\tfrac{1}{2}s + 1) &= \Im \log\Gamma(\tfrac{1}{2}iT + \tfrac{5}{4}) \\ &= \Im[(\tfrac{1}{2}iT + \tfrac{3}{4})\log(\tfrac{1}{2}iT + \tfrac{5}{4}) - \tfrac{1}{2}iT - \tfrac{5}{4} \\ &\qquad + \tfrac{1}{2}\log 2\pi + O(T^{-1})] \\ &= \tfrac{1}{2}T\log\tfrac{1}{2}T - \tfrac{1}{2}T + \tfrac{3}{8}\pi + O(T^{-1}).\end{aligned}$$

Hence

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + \frac{7}{8} + S(T) + O(T^{-1}), \tag{1}$$

where

$$\pi S(T) = \Delta_L \arg\zeta(s).$$

To prove (1) of §8, it suffices to prove that

$$S(T) = O(\log T). \tag{2}$$

This is one of the few estimates connected with $\zeta(s)$ that has not, as far as I know, been improved upon during the present century. Since $\arg\zeta(2) = 0$, we can express the definition of $S(T)$ in the form

$$\pi S(T) = \arg\zeta(\tfrac{1}{2} + iT),$$

provided this argument is defined by continuous variation along L, or, equivalently, by continuous horizontal movement from $+\infty + iT$ to $\frac{1}{2} + iT$, starting with the value 0. In view of our limited knowledge about $S(T)$, it would seem at first sight that we might as well omit the term $\frac{7}{8}$ in (1); but as we shall see later, it has a certain significance.

We shall base the proof of (2) on the following

Lemma. *If $\rho = \beta + i\gamma$ runs through the nontrivial zeros of $\zeta(s)$, then for large T*

$$\sum_\rho \frac{1}{1 + (T - \gamma)^2} = O(\log T). \tag{3}$$

For the proof, we refer to (4) of §13, which states that

$$-\Re\frac{\zeta'(s)}{\zeta(s)} < A\log t - \sum_\rho \Re\left(\frac{1}{s-\rho} + \frac{1}{\rho}\right)$$

for $1 \le \sigma \le 2$ and $t \ge 2$. In this formula we take $s = 2 + iT$.

Since $|\zeta'/\zeta|$ is bounded for such s, we obtain

$$\sum_\rho \Re\left(\frac{1}{s-\rho}+\frac{1}{\rho}\right) < A\log T.$$

As we have seen earlier, all the terms in both series are positive, and since

$$\Re\frac{1}{s-\rho}=\frac{2-\beta}{(2-\beta)^2+(T-\gamma)^2}\geq\frac{1}{4+(T-\gamma)^2},$$

we obtain the assertion in the lemma.

Two immediate corollaries are: (a) The number of zeros with $T-1<\gamma<T+1$ is $O(\log T)$; (b) the sum $\Sigma(T-\gamma)^{-2}$ extended over the zeros with γ outside the interval just mentioned is also $O(\log T)$.

Another deduction is that *for large t (not coinciding with the ordinate of a zero) and $-1\leq\sigma\leq 2$,*

$$\text{(4)}\qquad \frac{\zeta'(s)}{\zeta(s)}=\sum_\rho{}'\frac{1}{s-\rho}+O(\log t),$$

where the sum is limited to those ρ for which $|t-\gamma|<1$. For by (8) of §12, applied at s and at $2+it$ and subtracted,

$$\frac{\zeta'(s)}{\zeta(s)}=O(\log t)+\sum_\rho\left(\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\right).$$

For the terms with $|\gamma-t|\geq 1$, we have

$$\left|\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\right|=\frac{2-\sigma}{|(s-\rho)(2+it-\rho)|}\leq\frac{3}{|\gamma-t|^2},$$

and the sum of these is $O(\log t)$ by (b) above. As for the terms with $|\gamma-t|<1$, we have $|2+it-\rho|\geq 1$, and the number of terms is $O(\log t)$ by (a) above. Hence the result.

The estimate (2) for $S(T)$ follows easily from (4). For the definition of $S(T)$ implies that

$$\pi S(T)=O(1)-\int_{\frac{1}{2}+iT}^{2+iT}\Im[\zeta'(s)/\zeta(s)]\,ds,$$

the $O(1)$ term coming from the variation along $\sigma=2$. Now

$$\int_{\frac{1}{2}+iT}^{2+iT}\Im(s-\rho)^{-1}\,ds=\Delta\arg(s-\rho),$$

and this has absolute value at most π. The number of terms in the sum in (4) is $O(\log T)$ by (a) above, and therefore (2) follows.[1]

We have now proved the approximate formula for $N(T)$, from which it follows, incidentally, that if the ordinates $\gamma > 0$ are enumerated in increasing order as $\gamma_1, \gamma_2, \ldots$ then

$$\gamma_n \sim 2\pi n/\log n \qquad \text{as } n \to \infty.$$

It does not follow that $\gamma_{n+1} - \gamma_n \to 0$, but this result was proved by Littlewood in 1924. The formula for $N(T)$ shows that

$$N(T + H) - N(T) > A \log T \qquad (T > T_0)$$

if H is greater than some positive constant, and Titchmarsh proved the more precise result that this holds for any fixed positive H, with some positive A that depends on H. It may be noted that, in consequence, the estimate $O(\log T)$ in corollary (a) to the lemma is best possible.

As regards the function $S(T)$, it was proved by Littlewood that

$$\int_0^T S(t)\, dt = O(\log T),$$

and this indicates a high degree of cancellation among the values of the function. The result just stated would, of course, become false if the term $\frac{7}{8}$ had not been retained in (1).

For proofs of the results just stated, and for other results relative to the zeros, see Titchmarsh, Chap. 9.

[1] For another proof, see Titchmarsh, §9.4.

16

THE NUMBER $N(T, \chi)$

Let χ be a primitive character to the modulus q, and let $N(T, \chi)$ denote the number of zeros of $L(s, \chi)$ in the rectangle

$$0 < \sigma < 1, \qquad |t| < T.$$

(It is no longer appropriate to consider only the upper half-plane, since the zeros are not in general symmetrically placed with respect to the real axis.) In the present section we prove the approximate formula for $N(T, \chi)$ which corresponds to that for $N(T)$ proved in the preceding section. Since we regard $N(T, \chi)$ as a function of the two parameters T and q, it is no longer appropriate to suppose T arbitrarily large, and we merely assume that $T \geq 2$. The formula is

$$(1) \qquad \tfrac{1}{2}N(T, \chi) = \frac{T}{2\pi}\log\frac{qT}{2\pi} - \frac{T}{2\pi} + O(\log T + \log q).$$

[I have inserted a factor $\frac{1}{2}$ on the left for ease of comparison with $N(T)$, and to compensate for the rectangle being doubled.]

The proof is on the same lines as for $N(T)$. But it is convenient now to consider the variation in $\arg \xi(s, \chi)$ as s describes the rectangle R with vertices at

$$\tfrac{5}{2} - iT, \quad \tfrac{5}{2} + iT, \quad -\tfrac{3}{2} + iT, \quad -\tfrac{3}{2} - iT,$$

so as to avoid the possible zero at $s = -1$. This rectangle includes just one trivial zero of $L(s, \chi)$, at either $s = 0$ or $s = -1$, and therefore

$$2\pi[N(T, \chi) + 1] = \Delta_R \arg \xi(s, \chi).$$

The contribution of the left half of the contour is equal to that of the right half, since

$$\arg \xi(\sigma + it, \chi) = \arg \overline{\xi(1 - \sigma + it, \chi)} + c,$$

where c is independent of s.

By the definition of $\xi(s, \chi)$ in (12) of §12, we have to form the sum of

$$\Delta \arg(q/\pi)^{\frac{1}{2}s+\frac{1}{2}\mathfrak{a}} = T\log(q/\pi),$$
$$\Delta \arg \Gamma(\tfrac{1}{2}s + \tfrac{1}{2}\mathfrak{a}) = T\log\tfrac{1}{2}T - T + O(1),$$

and $\Delta \arg L(s, \chi)$, and then multiply by 2. The terms above give the main terms in (1), and it remains to prove, in effect, that

$$\arg L(\tfrac{1}{2} + iT, \chi) = O(\log T + \log q). \tag{2}$$

This follows, as before, from the following modified

Lemma. *If $\rho = \beta + i\gamma$ runs through the nontrivial zeros of $L(s, \chi)$, where χ is primitive, then for any real t,*

$$\sum_{\rho} \frac{1}{1 + (t - \gamma)^2} = O(\mathscr{L}), \tag{3}$$

where $\mathscr{L} = \log q(|t| + 2)$.

The proof is as before, but the reference is now to (2) of §14.

As in the preceding section, it follows from this lemma, in conjunction with (17) of §12, that *for t not coinciding with the ordinate of a zero, and $-1 \leq \sigma \leq 2$,*

$$\frac{L'(s, \chi)}{L(s, \chi)} = \sum_{\rho}{}' \frac{1}{s - \rho} + O(\mathscr{L}), \tag{4}$$

where the sum is limited to those ρ for which $|t - \gamma| < 1$.

The approximate formula (1) implies, in particular, that for large q the number of zeros with $|t| < T_0$, where T_0 is a suitable absolute constant, is greater than a constant multiple of $\log q$. This shows that the estimate

$$\sum_{\rho} \frac{1}{1 + \gamma^2} = O(\log q)$$

is essentially the best possible.[1]

For some purposes it is convenient to have an analog of (1) for characters that are not necessarily primitive. If χ is an imprimitive character, induced by the primitive character $\chi_1 (\mathrm{mod}\, q_1)$, then (1) remains valid for $N(T, \chi)$ as defined, provided we replace q by q_1.

[1] For some results on the zeros of $L(s, \chi)$ when q is large and t is bounded, see Siegel. *Annals of Math.*, **46**, 409–422 (1945), or *Gesammelte Abhandlungen* III, 47–60.

But if $N_R(T, \chi)$ denotes the number of zeros in the rectangle R defined above, we must include the zeros on $\sigma = 0$ of

$$\prod_{p|q} [1 - \chi_1(p)p^{-s}],$$

in accordance with (3) of §5. These are (for each p not dividing q_1) spaced at equal distances $2\pi/\log p$ apart. Their number, with $|t| < T$, is

$$O\left[\sum_{p|q} (T \log p + 1)\right] = O(T \log q).$$

Hence

$$N_R(T, \chi) = \frac{T}{\pi} \log \frac{T}{2\pi} + O(T \log q) \qquad \text{for } T \geq 2. \tag{5}$$

17

THE EXPLICIT FORMULA FOR $\psi(x)$

In this section we shall prove von Mangoldt's formula for $\psi(x)$, which was stated in §8. We recall that

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{p^m \leq x} \log p.$$

This function has discontinuities at the points where x is a prime power, and in order that the formula may remain valid at these points, it is necessary to modify the definition by taking the mean of the values on the left and on the right. In other words, we define $\psi_0(x)$ to be $\psi(x)$ when x is not a prime power, and $\psi(x) - \frac{1}{2}\Lambda(x)$ when it is. The formula asserts that, for $x > 1$,

$$\psi_0(x) = x - \sum_\rho \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log(1 - x^{-2}), \tag{1}$$

where the sum over the nontrivial zeros ρ of $\zeta(s)$ is to be understood in the symmetric sense as

$$\lim_{T \to \infty} \sum_{|\gamma| < T} \frac{x^\rho}{\rho}.$$

The value of the constant $\zeta'(0)/\zeta(0)$ is $\log 2\pi$, as can be deduced from (8) and (10) of §12. The last term of the formula is equivalent to $-\Sigma_\omega x^\omega/\omega$ extended over the trivial zeros of $\zeta(s)$ given by $\omega = -2, -4, -6, \ldots.$

To avoid some minor complications we shall suppose that $x \geq 2$, though as stated above the formula is valid for $x > 1$.

The general lines on which such a formula can be proved, provided that the argument can be justified, were indicated by Riemann in connection with his explicit formula for $\pi(x)$. The basic idea is to use the discontinuous integral

$$\frac{1}{2\pi i}\int_{c - i\infty}^{c + i\infty} y^s \frac{ds}{s} = \begin{cases} 0 & \text{if } 0 < y < 1, \\ \frac{1}{2} & \text{if } y = 1, \\ 1 & \text{if } y > 1, \end{cases} \tag{2}$$

where $c > 0$, to pick out the terms in a Dirichlet series with $n \le x$, by taking $y = x/n$. Since

$$\sum_{n=1}^{\infty} \Lambda(n)n^{-s} = -\zeta'(s)/\zeta(s)$$

for $\sigma > 1$, the result takes the form

$$\psi_0(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \left[-\frac{\zeta'(s)}{\zeta(s)}\right]\frac{x^s}{s}\,ds$$

for $c > 1$. If we can move the vertical line of integration away to infinity on the left, we shall express $\psi_0(x)$ as the sum of the residues of the function $[-\zeta'(s)/\zeta(s)]x^s/s$ at its poles. The pole of $\zeta(s)$ at $s = 1$ contributes x; the pole of $1/s$ at $s = 0$ contributes $-\zeta'(0)/\zeta(0)$; and each zero ρ of $\zeta(s)$, whether trivial or not, contributes $-x^\rho/\rho$.

To carry out this proof, we have to start with an integral from $c - iT$ to $c + iT$, and regard this as one side of a rectangle extending to the left. It is necessary to choose T with a little care, so that the horizontal sides of the rectangle shall avoid, as far as possible, the zeros of $\zeta(s)$ in the critical strip. After the argument has been carried out in detail, we shall have a finite form of (1), with an explicit estimate for the error; and this will be much more useful than (1) itself.

As a first step we prove the following

Lemma. *Let $\delta(y)$ denote the function of y on the right of* (2), *and let*

$$I(y, T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{y^s}{s}\,ds.$$

Then, for $y > 0, c > 0, T > 0$,

$$|I(y, T) - \delta(y)| < \begin{cases} y^c \min(1, T^{-1}|\log y|^{-1}) & \textit{if } y \ne 1, \\ cT^{-1} & \textit{if } y = 1. \end{cases}$$

Proof. Suppose first that $0 < y < 1$. The function y^s/s tends to 0 as $\sigma \to +\infty$, and does so uniformly in t. Hence we can replace the vertical integral by two horizontal integrals:

$$I(y, T) = -\frac{1}{2\pi i}\int_{c+iT}^{\infty+iT} \frac{y^s}{s}\,ds + \frac{1}{2\pi i}\int_{c-iT}^{\infty-iT} \frac{y^s}{s}\,ds.$$

Now

$$\left|\int_{c+iT}^{\infty+iT} \frac{y^s}{s}\,ds\right| \le \frac{1}{T}\int_c^{\infty} y^\sigma\,d\sigma = \frac{y^c}{T|\log y|},$$

and similarly for the other integral. This proves one of the two inequalities. The other is most easily obtained by replacing the vertical path by a circular path with center O, on the right side. The radius is $R = (c^2 + T^2)^{\frac{1}{2}}$, and on the circular arc we have $|y^s| \leq y^c$ and $|s| = R$. Hence

$$|I(y, T)| \leq \frac{1}{2\pi}\pi R \frac{y^c}{R} < y^c.$$

The proof when $y > 1$ is similar but uses a rectangle or circular arc to the left. The contour then includes the pole at $s = 0$, where the residue is $1 = \delta(y)$.

There remains the case $y = 1$, which is easily treated by direct computation. With $s = c + it$, we have

$$I(1, T) = \frac{1}{2\pi}\int_0^T \frac{2c}{c^2 + t^2}\,dt = \frac{1}{\pi}\int_0^{T/c} \frac{du}{1 + u^2}$$

$$= \frac{1}{2} - \frac{1}{\pi}\int_{T/c}^{\infty} \frac{du}{1 + u^2},$$

and the last integral is less than c/T. This proves the lemma.[1]

Applied to $\psi_0(x)$, the result of the lemma gives

$$(3)\qquad |\psi_0(x) - J(x, T)| < \sum_{\substack{n=1 \\ n \neq x}}^{\infty} \Lambda(n)(x/n)^c \min(1, T^{-1}|\log x/n|^{-1}) + cT^{-1}\Lambda(x),$$

where $c > 1$ and

$$(4)\qquad J(x, T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \left[-\frac{\zeta'(s)}{\zeta(s)}\right]\frac{x^s}{s}\,ds.$$

It is to be understood that the term containing $\Lambda(x)$ is present only if x is a prime power.

We choose $c = 1 + (\log x)^{-1}$, since this gives a good result without excessive work, and note that $x^c = ex$. We have to estimate the series on the right of (3), and we take first all terms for which

[1] It is an interesting exercise to prove the results for $y < 1$ and $y > 1$ by real variable methods.

$n \le \frac{3}{4}x$ or $n \ge \frac{5}{4}x$. For these, $|\log x/n|$ has a positive lower bound, and so their contribution to the sum is[2]

$$\ll xT^{-1} \sum_{n=1}^{\infty} \Lambda(n)n^{-c} = xT^{-1}\left[-\frac{\zeta'(c)}{\zeta(c)}\right] \ll xT^{-1}(\log x).$$

Consider next the terms for which $\frac{3}{4}x < n < x$. Let x_1 be the largest prime power less than x; we can suppose that $\frac{3}{4}x < x_1 < x$, since otherwise the terms under consideration vanish. For the term $n = x_1$, we have

$$\log\frac{x}{n} = -\log\left(1 - \frac{x - x_1}{x}\right) \ge \frac{x - x_1}{x},$$

and therefore the contribution of this term is

$$\ll \Lambda(x_1)\min\left[1, \frac{x}{T(x - x_1)}\right] \ll (\log x)\min\left[1, \frac{x}{T(x - x_1)}\right].$$

For the other terms, we can put $n = x_1 - \nu$, where $0 < \nu < \frac{1}{4}x$, and then

$$\log\frac{x}{n} \ge \log\frac{x_1}{n} = -\log\left(1 - \frac{\nu}{x_1}\right) \ge \frac{\nu}{x_1}.$$

Hence the contribution of these terms is

$$\ll \sum_{0<\nu<\frac{1}{4}x} \Lambda(x_1 - \nu)T^{-1}x_1/\nu \ll xT^{-1}(\log x)^2.$$

The terms with $x < n < \frac{5}{4}x$ are dealt with similarly, except that x_1 is replaced by x_2, the least prime power greater than x.

It is convenient to write $\langle x\rangle$ for the distance from x to the nearest prime power, other than x itself in case x is a prime power. Collecting the estimates, we deduce from (3) that

$$(5) \qquad |\psi_0(x) - J(x, T)| \ll \frac{x(\log x)^2}{T} + (\log x)\min\left(1, \frac{x}{T\langle x\rangle}\right).$$

The next step is to replace the vertical line of integration in (4) by the other three sides of the rectangle with vertices at

$$c - iT, \quad c + iT, \quad -U + iT, \quad -U - iT,$$

[2] From now on, we make use of Vinogradov's symbolism $A \ll B$, as an equivalent for $A = O(B)$.

where U is a large odd integer. Thus the left vertical side passes halfway between two of the trivial zeros of $\zeta(s)$. The sum of the residues of the integrand at its poles inside the rectangle is

$$x - \sum_{|\gamma|<T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \sum_{0<2m<U} \frac{x^{-2m}}{-2m}. \tag{6}$$

The choice of T demands consideration. We saw in a corollary to the lemma of §15 that, for any large T, the number of zeros with $|\gamma - T| < 1$ is $\ll \log T$. Among the ordinates of these zeros there must be a gap of length $\gg (\log T)^{-1}$. Hence by varying T by a bounded amount, we can ensure that

$$|\gamma - T| \gg (\log T)^{-1}$$

for all the zeros $\beta + i\gamma$.

We recall further the result of §15 that

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{|\gamma - T|<1} \frac{1}{s-\rho} + O(\log T)$$

for $s = \sigma + iT$ and $-1 \le \sigma \le 2$. With the present choice of T, each term is $\ll \log T$, and the number of terms is also $\ll \log T$; so that on the new horizontal lines of integration we have

$$\frac{\zeta'(s)}{\zeta(s)} = O(\log^2 T) \qquad \text{for } -1 \le \sigma \le 2.$$

The contribution made to the horizontal integrals by this range of σ is therefore

$$\ll \log^2 T \int_{-1}^{c} \left|\frac{x^s}{s}\right| d\sigma \ll \frac{\log^2 T}{T} \int_{-\infty}^{c} x^\sigma \, d\sigma \ll \frac{x \log^2 T}{T \log x}. \tag{7}$$

It remains to estimate the contribution made by the horizontal lines of integration for $-U \le \sigma \le -1$ and by the vertical line $\sigma = -U$. We need an estimate for $|\zeta'/\zeta|$ for $\sigma \le -1$, and we shall prove that

$$|\zeta'(s)/\zeta(s)| \ll \log(2|s|) \tag{8}$$

in this half-plane, provided that circles of radius $\frac{1}{2}$ (say) around all the trivial zeros at $s = -2, -4, \ldots$ are excluded. It will follow that the contribution of the remainder of the horizontal integrals is

$$\ll \frac{\log 2T}{T} \int_{-U}^{-1} x^\sigma \, d\sigma \ll \frac{\log T}{Tx \log x},$$

which is negligible compared with (7), and the contribution of the vertical integral is

$$\ll \frac{\log 2U}{U}\int_{-T}^{T} x^{-U}\,dt \ll \frac{T\log U}{Ux^{U}},$$

which vanishes as $U \to \infty$.

Adding the estimate in (7) to that in (5), and making $U \to \infty$ in (6), we obtain

(9) $$\psi_0(x) = x - \sum_{|\gamma|<T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \tfrac{1}{2}\log(1 - x^{-2}) + R(x, T),$$

where

(10) $$|R(x, T)| \ll \frac{x\log^2(xT)}{T} + (\log x)\min\left(1, \frac{x}{T\langle x\rangle}\right).$$

The estimate (8) is deduced from the functional equation, which is best taken in its unsymmetric form [(4) of §10]

$$\zeta(1 - s) = 2^{1-s}\pi^{-s}(\cos\tfrac{1}{2}\pi s)\Gamma(s)\zeta(s),$$

since, if $1 - \sigma \leq -1$ the functions on the right have to be considered only for $\sigma \geq 2$. The logarithmic derivative of the right side, apart from an added constant, is

$$-\tfrac{1}{2}\pi\tan\tfrac{1}{2}\pi s + \frac{\Gamma'(s)}{\Gamma(s)} + \frac{\zeta'(s)}{\zeta(s)}.$$

The first term is bounded if $|s - (2m + 1)| \geq \frac{1}{2}$, that is, if

$$|(1 - s) + 2m| \geq \tfrac{1}{2}.$$

The second term is $\ll \log|s|$, and therefore $\ll \log 2|1 - s|$ for $\sigma \geq 2$. The third term is bounded. Hence (8) follows.

The results (9) and (10) constitute the more precise form of the explicit formula (1). As $T \to \infty$ for any given $x \geq 2$, we have $R(x, T) \to 0$, and therefore (1) follows. The convergence is uniform in any closed interval of x which does not contain a prime power, but not otherwise, since $\psi_0(x)$ is discontinuous at each prime power value of x.

We proved (9) and (10) subject to a restriction on T, but this can now be removed. The effect of varying T by a bounded amount is to change the sum over ρ by $O(\log T)$ terms, and each term is $O(x/T)$. Hence the variation in the sum is $O[x(\log T)/T]$, and this is covered by the estimate on the right of (10).

We note for future reference that, if x is an integer, then $\langle x \rangle \geq 1$, and (10) takes the simpler form

$$|R(x, T)| \ll x(\log xT)^2 T^{-1}. \tag{11}$$

The results (9) and (10) continue to hold[3] for $1 < x < 2$, with a slight modification in the form of the estimate for $R(x, T)$.

[3] For this, and for a discussion of the series $\Sigma x^\rho/\rho$ when $0 < x < 1$, see Ingham, p. 81.

18

THE PRIME NUMBER THEOREM

We shall now deduce, from the results of the last section and those of §13, that

(1) $$\psi(x) = x + O\{x \exp[-c(\log x)^{\frac{1}{2}}]\},$$

and from this the analogous result for $\pi(x)$, which includes the prime number theorem. This is by no means the easiest way of proving the prime number theorem, but it is an instructive way. It is also very close to the method used by de la Vallée Poussin, though he worked with the function

$$\psi_1(x) = \sum_{n \leq x} (x - n)\Lambda(n)$$

instead of the function $\psi(x)$.

The main question is that of estimating the sum $\Sigma x^\rho/\rho$ in (9) of the preceding section, and obviously any effective estimate must be deduced from the fact that the real part β of ρ is not too near 1. It follows from the result of §13 that if $|\gamma| < T$, where T is large, then $\beta < 1 - c_1/\log T$, where c_1 is a positive absolute constant. Hence

$$|x^\rho| = x^\beta < x \exp[-c_1(\log x)/(\log T)].$$

Also $|\rho| \geq \gamma$, for $\gamma > 0$; so it remains to estimate

$$\sum_{0<\gamma<T} 1/\gamma.$$

If $N(t)$ denotes, as in §15, the number of zeros in the critical strip with ordinates less than t, this sum is

$$\int_0^T t^{-1}\, dN(t) = \frac{1}{T} N(T) + \int_0^T t^{-2} N(t)\, dt.$$

Since $N(t) \ll t \log t$ for large t, this is $\ll \log^2 T$. Hence

$$\sum_{|\gamma|<T} \left|\frac{x^\rho}{\rho}\right| \ll x(\log T)^2 \exp[-c_1(\log x)/(\log T)].$$

We can take x to be an integer, without loss of generality. It follows from (9) and (11) of §17 that

$$|\psi(x) - x| \ll \frac{x \log^2(xT)}{T} + x(\log T)^2 \exp[-c_1(\log x)/(\log T)],$$

for large x. If we determine T as a function of x by

$$(\log T)^2 = \log x,$$

so that

$$T^{-1} = \exp[-(\log x)^{\frac{1}{2}}],$$

we get

$$\begin{aligned}|\psi(x) - x| &\ll x(\log x)^2 \exp[-(\log x)^{\frac{1}{2}}] + x(\log x) \exp[-c_1(\log x)^{\frac{1}{2}}] \\ &\ll x \exp[-c_2(\log x)^{\frac{1}{2}}],\end{aligned}$$

provided c_2 is a constant that is less than both 1 and c_1. This proves (1).

The transition to an asymptotic formula for $\pi(x)$, instead of for $\psi(x)$, is elementary and is essentially an exercise in partial summation. First we pass to the function

$$\pi_1(x) = \sum_{n \le x} \frac{\Lambda(n)}{\log n}.$$

This is expressed in terms of the function $\psi(x)$ by

$$\begin{aligned}\pi_1(x) &= \sum_{n \le x} \Lambda(n) \int_n^x \frac{dt}{t \log^2 t} + \frac{1}{\log x} \sum_{n \le x} \Lambda(n) \\ &= \int_2^x \frac{\psi(t)\, dt}{t \log^2 t} + \frac{\psi(x)}{\log x}.\end{aligned}$$

The effect of replacing $\psi(t)$ by t is to give

$$\int_2^x t \frac{d}{dt}\left(-\frac{1}{\log t}\right) dt + \frac{x}{\log x} = \mathrm{li}\, x + \frac{2}{\log 2},$$

on integrating by parts. Thus it remains to consider the estimate for the error term, which is

$$\ll \int_2^x \exp[-c_2(\log t)^{\frac{1}{2}}]\, dt + x \exp[-c_2(\log x)^{\frac{1}{2}}].$$

The contribution of the range $t < x^{\frac{1}{4}}$ to the integral is trivially less than $x^{\frac{1}{4}}$, and in the rest of the range we have $(\log t)^{\frac{1}{2}} > \frac{1}{2}(\log x)^{\frac{1}{2}}$.

Hence

$$\pi_1(x) = \text{li}\, x + O\{x \exp[-c_3(\log x)^{\frac{1}{2}}]\},$$

where $c_3 = \frac{1}{2}c_2$.

Finally, since

$$\pi_1(x) = \sum_{p^m \le x} \frac{\log p}{m \log p}$$
$$= \pi(x) + \tfrac{1}{2}\pi(x^{\frac{1}{2}}) + \tfrac{1}{3}\pi(x^{\frac{1}{3}}) + \cdots,$$

and $\pi(x^{\frac{1}{2}}) \le x^{\frac{1}{2}}$, $\pi(x^{\frac{1}{3}}) \le x^{\frac{1}{3}}, \ldots$, the difference between $\pi_1(x)$ and $\pi(x)$ is $O(x^{\frac{1}{2}})$. Thus

$$\pi(x) = \text{li}\, x + O\{x \exp[-c_3(\log x)^{\frac{1}{2}}]\}.$$

This is the form of the prime number theorem proved by de la Vallée Poussin in 1899. It was improved to

$$\pi(x) = \text{li}\, x + O\{x \exp[-c(\theta)(\log x)^{\theta}]\}$$

for any $\theta < \frac{3}{5}$, by Vinogradov and Korobov in 1958. The improvement comes from the result on a zero-free region for $\zeta(s)$, mentioned at the end of §13. One uses this in the explicit formula for $\psi(x)$, and chooses T so that $(\log T)^{1+\alpha} = \log x$, where α is any number greater than $\frac{2}{3}$.

The assumption of the Riemann hypothesis implies a much better estimate for the error term, as was pointed out by von Koch in 1901. We then have $|x^{\rho}| = x^{\frac{1}{2}}$, and as we proved earlier that $\Sigma 1/|\rho|$ over $0 < \gamma < T$ is $O(\log^2 T)$, the explicit formula gives

$$|\psi(x) - x| \ll x^{\frac{1}{2}} \log^2 T + xT^{-1} \log^2 xT,$$

if x is an integer. Choosing $T = x^{\frac{1}{2}}$, we get

$$\psi(x) = x + O(x^{\frac{1}{2}} \log^2 x).$$

From this it follows, by the same argument as above, that

$$\pi(x) = \text{li}\, x + O(x^{\frac{1}{2}} \log x).$$

The situation is exactly similar, with x^{Θ} in place of $x^{\frac{1}{2}}$, if one assumes only that all the zeros have $\beta \le \Theta$, where Θ is a number between $\frac{1}{2}$ and 1.

There is also an implication in the opposite sense, by an argument which is quite elementary. If we assume that

$$\psi(x) = x + O(x^{\alpha})$$

for some fixed $\alpha < 1$, it follows that all the zeros of $\zeta(s)$ have $\beta \leq \alpha$. For if $\sigma > 1$, we have

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \Lambda(n) n^{-s},$$

and this is easily rearranged in the form

$$-\frac{\zeta'(s)}{\zeta(s)} = s \int_1^{\infty} \psi(x) x^{-s-1}\, dx,$$

as on similar occasions earlier. If $\psi(x) = x + R(x)$, we get

$$-\frac{\zeta'(s)}{\zeta(s)} = \frac{s}{s-1} + s \int_1^{\infty} R(x) x^{-s-1}\, dx.$$

The supposition that $R(x) = O(x^{\alpha})$ implies that the integral represents a regular function of s for $\sigma > \alpha$, and then $\zeta(s)$ can have no zeros in this half-plane.

There is the curious conclusion, from the last two results, that if $\psi(x) = x + O(x^{\Theta + \varepsilon})$ for each $\varepsilon > 0$, where Θ is a fixed number between $\frac{1}{2}$ and 1, then necessarily

$$\psi(x) = x + O(x^{\Theta} \log^2 x).$$

Grosswald[1] has shown that if Θ is strictly larger than $\frac{1}{2}$ then the factor $\log^2 x$ can be deleted.

[1] *C. R. Acad. Sci., Paris*, **260**, 3813–3816 (1965).

19

THE EXPLICIT FORMULA FOR $\psi(x, \chi)$

For any character χ to the modulus q, we define

$$\psi(x, \chi) = \sum_{n \leq x} \chi(n)\Lambda(n). \tag{1}$$

These sums play much the same part in the prime number theorem for arithmetic progressions as that played by $\psi(x)$ in the prime number theorem itself, but now there is an aggregate of $\phi(q)$ such sums, one for each character, instead of a single sum. In this section we shall establish the explicit formula that is analogous to that proved for $\psi(x)$ in §17. As there, we modify $\psi(x, \chi)$ to $\psi_0(x, \chi)$ in case x is a prime power.

The general lines of the argument are the same as in §17, but with L'/L in place of ζ'/ζ. Suppose χ is a primitive character (mod q). Consider first the computation of the residues of

$$-\frac{L'(s, \chi)}{L(s, \chi)}\frac{x^s}{s}.$$

The only poles are at the zeros of $L(s, \chi)$ and at $s = 0$; and there is a slight complication in that, if $\chi(-1) = 1$, one of the zeros of $L(s, \chi)$ is itself at $s = 0$, so that the function has a double pole there.

Suppose first that $\chi(-1) = -1$. Then the complication just mentioned does not arise, and the explicit formula is

$$\psi_0(x, \chi) = -\sum_\rho \frac{x^\rho}{\rho} - \frac{L'(0, \chi)}{L(0, \chi)} + \sum_{m=1}^{\infty} \frac{x^{1-2m}}{2m-1}, \tag{2}$$

the expression on the right being the sum of the residues of the function mentioned above. There is the same understanding about the sum over the nontrivial zeros ρ of $L(s, \chi)$ as in §17. The value of $L'(0, \chi)/L(0, \chi)$ can be expressed in terms of the constant $B(\chi)$ of §12 by putting $s = 0$ in (17) of §12.

Suppose next that $\chi(-1) = 1$. Since $L(s, \chi)$ has a simple zero at $s = 0$, the expansion near $s = 0$ of L'/L is of the form

$$\frac{L'(s, \chi)}{L(s, \chi)} = \frac{1}{s} + b + \cdots,$$

where $b = b(\chi)$. Since

$$\frac{x^s}{s} = \frac{1}{s} + (\log x) + \cdots,$$

the residue of the function mentioned above at $s = 0$ is $-(\log x + b)$. The explicit formula therefore takes the form

$$\psi_0(x, \chi) = -\sum_{\rho} \frac{x^{\rho}}{\rho} - \log x - b(\chi) + \sum_{m=1}^{\infty} \frac{x^{-2m}}{2m}. \tag{3}$$

Once again, $b(\chi)$ can be expressed in terms of $B(\chi)$ by using (17) of §12.

We now outline the proof, and the estimation of the error term when the sum is taken over $|\gamma| < T$. We suppose that $x \geq 2$ and $T \geq 2$. The character $\chi(n)$ plays no part in the estimation of the sum on the right of (3) of §17, and therefore the inequality analogous to (5) of §17 is still valid. In the choice of a modified value of T, we have to appeal to the lemma of §16 instead of that of §15, and accordingly we get

$$\frac{L'(\sigma \pm iT, \chi)}{L(\sigma \pm iT, \chi)} = O(\log^2 qT) \qquad \text{for } -1 \leq \sigma \leq 2.$$

The contribution made to the horizontal integrals by this range of σ is therefore

$$\ll \frac{x \log^2 qT}{T \log x}.$$

The estimate for L'/L in the half-plane $\sigma \leq -1$, when circles of radius $\frac{1}{2}$ around the trivial zeros are excluded, is

$$\frac{L'(s, \chi)}{L(s, \chi)} = O[\log(q|s|)].$$

This follows by logarithmic differentiation from the functional equation of $L(s, \chi)$ in its unsymmetric form, namely,

$$L(1 - s, \chi) = \varepsilon(\chi) 2^{1-s} \pi^{-s} q^{s-\frac{1}{2}} \cos \tfrac{1}{2}\pi(s - \mathfrak{a}) \Gamma(s) L(s, \bar{\chi}),$$

where $|\varepsilon(\chi)| = 1$ and $\mathfrak{a} = 0$ or 1. This form is deduced from the symmetric form in the same way as (4) of §10. The contribution of the rest of the horizontal integrals is therefore

$$\ll \frac{\log qT}{Tx \log x},$$

and as before it is negligible.

The result is that

$$(4) \qquad \psi_0(x, \chi) = - \sum_{|\gamma| < T} \frac{x^\rho}{\rho} - (1 - \mathfrak{a}) \log x - b(\chi) + \sum_{m=1}^{\infty} \frac{x^{\mathfrak{a} - 2m}}{2m - \mathfrak{a}} + R(x, T),$$

where

$$(5) \qquad |R(x, T)| \ll \frac{x}{T} \log^2 qxT + (\log x) \min\left(1, \frac{x}{T\langle x \rangle}\right).$$

Again, if x is fixed, this tends to 0 as $T \to \infty$, and so we have the results given in (2) for the case $\mathfrak{a} = 1$ and in (3) for the case $\mathfrak{a} = 0$. If x is an integer, we can replace (5) by

$$(6) \qquad |R(x, T)| \ll xT^{-1} \log^2(qxT).$$

From the point of view of application to the distribution of primes in arithmetic progressions with a variable modulus, formula (4) is of little use as it stands. It contains the unknown $b(\chi)$, and it contains terms x^ρ/ρ for which ρ may be very near either 1 or 0. It will be recalled that the results of §14 state that there is at most one zero within a distance $c/\log q$ of $s = 1$ (and so also of $s = 0$), and this one zero can only occur when χ is a real character and is itself real. It is important to have this zero visible explicitly in the formula.

We need no longer distinguish between ψ and ψ_0, as we are not aiming at exactitude, and we can simplify (4) by absorbing $\log x$ and the sum over m into the error term. We can use the form (6) of the error term, since the effect on $\psi_0(x, \chi)$ of replacing x by the nearest integer is $O(\log x)$; and for simplicity we suppose now that $T \leq x$. Then

$$(7) \qquad \psi(x, \chi) = - \sum_{|\gamma| < T} \frac{x^\rho}{\rho} - b(\chi) + R_1(x, T),$$

$$(8) \qquad |R_1(x, T)| \ll xT^{-1} \log^2 qx.$$

The first step is to express $b(\chi)$ in another form. By (17) of §12,

$$\frac{L'(s,\chi)}{L(s,\chi)} = -\frac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})}{\Gamma(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})} + B(\chi) + \sum_\rho\left(\frac{1}{s-\rho}+\frac{1}{\rho}\right).$$

Replacing s by 2 and subtracting, we obtain

$$\frac{L'(s,\chi)}{L(s,\chi)} = O(1) - \frac{1}{2}\frac{\Gamma'(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})}{\Gamma(\frac{1}{2}s+\frac{1}{2}\mathfrak{a})} + \sum_\rho\left(\frac{1}{s-\rho}-\frac{1}{2-\rho}\right),$$

where the $O(1)$ is absolute. If $\chi(-1) = -1$, so that $\mathfrak{a} = 1$, the term Γ'/Γ is regular at $s = 0$; if $\chi(-1) = 1$, so that $\mathfrak{a} = 0$, its expansion near $s = 0$ is s^{-1} + const. $+\cdots$. Hence the number $b(\chi)$, which we defined earlier as the value of $L'(0,\chi)/L(0,\chi)$ in the former case, and as the constant term in the expansion of $L'(s,\chi)/L(s,\chi)$ near $s = 0$ in the latter case, satisfies

$$b(\chi) = O(1) - \sum_\rho\left(\frac{1}{\rho}+\frac{1}{2-\rho}\right).$$

For the terms in this series with $|\gamma| \geq 1$, we have

$$\sum_{|\gamma|\geq 1}\left|\frac{1}{\rho}+\frac{1}{2-\rho}\right| = 2\sum_{|\gamma|\geq 1}\frac{1}{|\rho(2-\rho)|} \ll \sum_\rho\frac{1}{|2-\rho|^2}.$$

The last sum can be estimated as $O(\log q)$ using (3) of §16 with $t = 0$. The same estimate applies to

$$\sum_{|\gamma|<1}\frac{1}{2-\rho},$$

since for $|\gamma| < 1$ we have $|2-\rho| \gg |2-\rho|^2$. It follows that

$$b(\chi) = O(\log q) - \sum_{|\gamma|<1}\frac{1}{\rho}.$$

We can therefore rewrite (7) as

$$(9)\qquad \psi(x,\chi) = -\sum_{|\gamma|<T}\frac{x^\rho}{\rho} + \sum_{|\gamma|<1}\frac{1}{\rho} + R_2(x,T),$$

where $R_2(x,T)$ satisfies (8).

By the theorem of §14, there is no zero of $L(s,\chi)$ satisfying

$$(10)\qquad |\gamma| < 1, \qquad \beta > 1 - c/\log q,$$

except possibly when χ is real, when there may (as far as we know) be one simple real zero. Here c is a numerical constant, which we

can suppose less than $\frac{1}{4}$, whence $\beta > \frac{3}{4}$, since $q \geq 3$. We call such a real zero *exceptional* and denote it by β_1. There will also be a zero at $1 - \beta_1$.

Let Σ' denote a summation over the zeros which excludes the possible zeros β_1 and $1 - \beta_1$. Then we can rewrite (9) as

$$\psi(x, \chi) = - \sum_{|\gamma| < T}' \frac{x^\rho}{\rho} + \sum_{|\gamma| < 1}' \frac{1}{\rho} - \frac{x^{\beta_1} - 1}{\beta_1} - \frac{x^{1-\beta_1} - 1}{1 - \beta_1} + R_2(x, T).$$

The second sum can be absorbed in the error term, since

$$\rho^{-1} = O(\log q)$$

for the zeros in question, and their number is $O(\log q)$ by (1) of §16 with $T = 2$. We can also omit the term β_1^{-1}, which is $O(1)$. Finally,

$$\frac{x^{1-\beta_1} - 1}{1 - \beta_1} = x^\sigma \log x$$

for some σ between 0 and $1 - \beta_1$, and the last expression is less than $x^{\frac{1}{4}} \log x$.

We now have the convenient expression (valid for primitive χ and $2 \leq T \leq x$)

$$\psi(x, \chi) = - \frac{x^{\beta_1}}{\beta_1} - \sum_{|\gamma| < T}' \frac{x^\rho}{\rho} + R_3(x, T), \tag{11}$$

where

$$|R_3(x, T)| \ll xT^{-1} \log^2(qx) + x^{\frac{1}{4}} \log x. \tag{12}$$

The term $-x^{\beta_1}/\beta_1$ in (11) can only occur if χ is real.

Finally, we prove that (11) and (12) hold for any nonprincipal character χ, whether primitive or not. Suppose χ is imprimitive and is induced by the primitive character $\chi_1 \pmod{q_1}$. The difference between $\psi(x, \chi)$ and $\psi(x, \chi_1)$ is at most

$$\sum_{\substack{n \leq x \\ (n,q) > 1}} \Lambda(n) = \sum_{p|q} \sum_{\substack{\nu \\ p^\nu \leq x}} \log p \ll (\log x) \sum_{p|q} \log p \ll (\log x)(\log q),$$

which is negligible compared with the expression in (12), where $T \leq x$. This expression applies to the error term in the formula for $\psi(x, \chi_1)$ because $q > q_1$.

There is, however, a logical point that needs consideration. The definition of the exceptional zero β_1 is a definition that involves the modulus, and, assuming we use the same definition when χ is imprimitive, an exceptional zero for $L(s, \chi)$ will certainly be an exceptional zero for $L(s, \chi_1)$, but not necessarily vice versa. However,

if a zero is exceptional for χ_1 but not for χ, the term $-x^{\beta_1}/\beta_1$, which is explicit in the formula for $\psi(x, \chi_1)$, will still be present in the formula for $\psi(x, \chi)$, since it will occur there in the sum $-\sum' x^\rho/\rho$.

Thus the formula remains valid, and we restate it for convenience of reference:

If χ is a nonprincipal character to the modulus q, and $2 \leq T \leq x$, then

(13) $$\psi(x, \chi) = -\frac{x^{\beta_1}}{\beta_1} - \sum_{|\gamma|<T}' \frac{x^\rho}{\rho} + R_3(x, T),$$

where

(14) $$|R_3(x, T)| \ll xT^{-1} \log^2 qx + x^{\frac{1}{4}} \log x.$$

The term $-x^{\beta_1}/\beta_1$ is to be omitted unless χ is a real character for which $L(s, \chi)$ has a zero β_1 (which is necessarily unique and simple) satisfying

(15) $$\beta_1 > 1 - c/\log q,$$

where c is a certain absolute constant; and the sum Σ' excludes β_1 and $1 - \beta_1$ (if they exist).

As we saw in §14 there is at most one real character (mod q) for which such a zero β_1 can exist. It may be noted that the term $x^{\frac{1}{4}} \log x$ in (14) can be omitted unless β_1 exists.

20

THE PRIME NUMBER THEOREM FOR ARITHMETIC PROGRESSIONS (I)

We now apply the last result of the preceding section to obtain approximations to

$$(1)\qquad \psi(x;q,a) = \sum_{\substack{n \le x \\ n \equiv a \,(\mathrm{mod}\, q)}} \Lambda(n).$$

From this it is an elementary matter to deduce approximations to $\pi(x;q,a)$, the number of primes up to x that are congruent to $a \,(\mathrm{mod}\, q)$.

The relationship between $\psi(x;q,a)$ and the sums $\psi(x,\chi)$ of the preceding section follows immediately from (4) of §4; we have

$$(2)\qquad \psi(x;q,a) = \frac{1}{\phi(q)} \sum_{\chi} \bar{\chi}(a)\psi(x,\chi),$$

where the sum is over all the characters χ to the modulus q.

The contribution of the principal character χ_0 to the sum on the right provides the main term. By an argument similar to one used in the preceding section, we have

$$|\psi(x,\chi_0) - \psi(x)| \le \sum_{\substack{n \le x \\ (n,q) > 1}} \Lambda(n) \ll (\log q)(\log x).$$

By de la Vallée Poussin's form of the prime number theorem, namely (1) of §18,

$$\psi(x) = x + O\{x \exp[-c_1(\log x)^{\frac{1}{2}}]\},$$

where c_1 is a positive constant. Hence

$$(3)\qquad \psi(x;q,a) = \frac{x}{\phi(q)} + \frac{1}{\phi(q)} \sum_{\chi \ne \chi_0} \bar{\chi}(a)\psi(x,\chi) + O\left(\frac{1}{\phi(q)}\{x \exp[-c_1(\log x)^{\frac{1}{2}}] + \log^2 qx\}\right).$$

For $\psi(x, \chi)$ when $\chi \neq \chi_0$, we have the expression (13) of the last section, namely

$$\psi(x, \chi) = -\frac{x^{\beta_1}}{\beta_1} - \sum_{|\gamma|<T}' \frac{x^\rho}{\rho} + R_3(x, T), \tag{4}$$

where

$$|R_3(x, T)| \ll xT^{-1}\log^2 qx + x^{\frac{1}{4}}\log x$$

provided $2 \leq T \leq x$. The term in (4) containing β_1 occurs for at most one real nonprincipal χ.

By the results of §14, and the remarks at the end of §19, all the zeros ρ in the sum on the right of (4) satisfy

$$\beta < 1 - c_2/\log qT$$

for a certain positive absolute constant c_2. Hence

$$|x^\rho| = x^\beta < x\exp[-c_2(\log x)/(\log qT)].$$

The sum $\Sigma|\rho|^{-1}$ extended over the zeros in (4) with $|\gamma| > 1$ can be estimated as in §18, and is

$$\ll \int_1^T t^{-2}N(t, \chi)\,dt \ll \int_1^T t^{-1}\log(qt)\,dt \ll \log^2 qT \quad \ll \log^2 qx.$$

The same sum over the zeros ρ with $|\gamma| \leq 1$ is $O(\log^2 q)$, since $|\rho|^{-1} = O(\log q)$ for each of them. Hence

$$\psi(x, \chi) = -\frac{x^{\beta_1}}{\beta_1} + R_4(x, T), \tag{5}$$

where

$$|R_4(x, T)| \ll x(\log^2 qx)\exp[-c_2(\log x)/(\log qT)] \tag{6}$$
$$+ xT^{-1}\log^2 qx + x^{\frac{1}{4}}\log x.$$

Some condition must be imposed on the size of q in relation to that of x. If we suppose that

$$q \leq \exp[C(\log x)^{\frac{1}{2}}], \tag{7}$$

where C is any positive constant, and choose

$$T = \exp[C(\log x)^{\frac{1}{2}}],$$

then all the terms on the right of (6) are

$$\ll x\exp[-C'(\log x)^{\frac{1}{2}}]$$

for some positive C' depending only on C. Hence, *subject to* (7), *we have*

$$(8) \qquad \psi(x, \chi) = -x^{\beta_1}/\beta_1 + O\{x \exp[-C'(\log x)^{\frac{1}{2}}]\}$$

for each nonprincipal χ *to the modulus* q.

Substituting in (3), and recalling that the term containing β_1 occurs for at most one χ, we get the following result for $\psi(x; q, a)$:

Let C be any positive constant. Then

$$(9) \qquad \psi(x; q, a) = \frac{x}{\phi(q)} - \frac{\bar{\chi}_1(a)x^{\beta_1}}{\phi(q)\beta_1} + O\{x \exp[-C'(\log x)^{\frac{1}{2}}]\}$$

for a positive constant C' *depending only on* C, *and this holds uniformly with respect to* q *in the range* (7). *Here* χ_1 *is the single real character* (*mod* q), *if it exists, for which* $L(s, \chi_1)$ *has a real zero* β_1 *satisfying* $\beta_1 > 1 - c/\log q$ *for a certain positive absolute constant* c.

It is in the possible term containing β_1 that one of the main difficulties in the theory of the distribution of primes in arithmetic progressions shows itself. The only universal upper bound that we have for β_1 is (12) of §14, which states that $\beta_1 < 1 - c_3/q^{\frac{1}{2}} \log^2 q$. The term containing β_1 is therefore

$$\ll \frac{x}{\phi(q)} \exp\left(-c_3 \frac{\log x}{q^{\frac{1}{2}} \log^2 q}\right).$$

This will only be of the same order as the other error term in (9) if we impose a very severe limitation on q, such as

$$(10) \qquad q \leq (\log x)^{1-\delta}$$

for some fixed $\delta > 0$. We then obtain the following result.

Provided q *satisfies* (10) *for some fixed* $\delta > 0$, *we have*

$$(11) \qquad \psi(x; q, a) = \frac{x}{\phi(q)} + O\{x \exp[-c_4(\log x)^{\frac{1}{2}}]\},$$

where c_4 *is an absolute constant.*

This is a weak result, but as far as I know it is as yet the only result of the kind (apart from minor variations) that is effective, in the sense that, if δ is given a numerical value, both c_4 and the constant implied by the symbol O can be given numerical values.

As Page showed, we can obtain a similar result in the wider range (7), provided q does not coincide with a multiple of a particular integer q_1 depending on x. In his result, given in §14, we take

$$z = \exp[C(\log x)^{\frac{1}{2}}].$$

Then the result tells us that there is at most one real primitive character to a modulus not exceeding z for which

$$\beta_1 > 1 - \frac{c_5}{\log z} = 1 - \frac{c_5}{C(\log x)^{\frac{1}{2}}}.$$

Denote the modulus of this character (if it exists) by q_1. Then if q is not a multiple of q_1, we have

$$\beta_1 \leq 1 - \frac{c_5}{C(\log x)^{\frac{1}{2}}}$$

for every real nonprincipal $\chi \pmod q$, whether primitive or not. We then obtain the same type of estimate for the term containing β_1 as before. Note that, if q_1 exists, it must satisfy

$$1 - \frac{c_5}{C(\log x)^{\frac{1}{2}}} \leq 1 - \frac{c_3}{q_1^{\frac{1}{2}} \log^2 q_1},$$

that is,

$$(12) \qquad q_1 \log^4 q_1 \gg \log x.$$

Hence we have proved:

Let C be any constant. Then, except possibly if q is a multiple of a particular integer q_1 depending on x, we have

$$(13) \qquad \psi(x; q, a) = \frac{x}{\phi(q)} + O\{x \exp[-C''(\log x)^{\frac{1}{2}}]\}$$

for a positive constant C″ depending only on C, and this holds uniformly with respect to q in the range (7). *The integer q_1 satisfies* (12).

In the next section we shall prove Siegel's theorem, which gives a much better upper bound for β_1 than we have had hitherto, and then in §22 we shall return to the question of the distribution of primes in arithmetic progressions.

We conclude this section by stating the consequences of the generalized Riemann hypothesis, that is, the hypothesis that not only $\zeta(s)$ but all the functions $L(s, \chi)$ have their zeros in the critical strip on the line $\sigma = \frac{1}{2}$. (This conjecture seems to have been first formulated by Piltz in 1884.) Then

$$\psi(x) = x + O(x^{\frac{1}{2}} \log^2 x),$$

as we saw in §18, and the same holds for $\psi(x, \chi_0)$, by the inequality for $|\psi(x, \chi_0) - \psi(x)|$ at the beginning of this section, provided we

suppose (say) that $q \leq x$. When $\chi \neq \chi_0$, (13) of §19 implies that, on the above hypothesis,

$$|\psi(x, \chi)| \ll x^{\frac{1}{2}} + x^{\frac{1}{2}} \sum_{|\gamma| < T} |\rho|^{-1} + xT^{-1} \log^2 qx + x^{\frac{1}{4}} \log x$$

for $2 \leq T \leq x$. As proved earlier in this section,

$$\sum |\rho|^{-1} \ll \log^2 qx.$$

Taking $T = x^{\frac{1}{2}}$, we get

$$|\psi(x, \chi)| \ll x^{\frac{1}{2}} \log^2 x$$

for $\chi \neq \chi_0$ and $q \leq x$. It now follows from (2) that *on the generalized Riemann hypothesis, if $q \leq x$,*

$$\psi(x; q, a) = \frac{x}{\phi(q)} + O(x^{\frac{1}{2}} \log^2 x). \tag{14}$$

It will be seen that even this powerful hypothesis gives only a poor result if q is larger than $x^{\frac{1}{2}}$.

21

SIEGEL'S THEOREM

Siegel's theorem,[1] in the first of its two forms, states that:

For any $\varepsilon > 0$ *there exists a positive number* $C_1(\varepsilon)$ *such that, if* χ *is a real primitive character to the modulus* q, *then*

$$L(1, \chi) > C_1(\varepsilon)q^{-\varepsilon}. \tag{1}$$

This implies, by (15) and (16) of §6, that

$$h(d) > C_2(\varepsilon)|d|^{\frac{1}{2}-\varepsilon} \qquad \text{for } d < 0 \tag{2}$$

and

$$h(d)\log \eta > C_2(\varepsilon)d^{\frac{1}{2}-\varepsilon} \qquad \text{for } d > 0, \tag{3}$$

where $\eta = \frac{1}{2}(t_0 + u_0\sqrt{d})$ and t_0, u_0 have the same meaning as in §6.

In its second form, the theorem states that:

For any $\varepsilon > 0$ *there exists a positive number* $C_3(\varepsilon)$ *such that, if* χ *is any real nonprincipal character, with modulus* q, *then* $L(s, \chi) \neq 0$ *for*

$$s > 1 - C_3(\varepsilon)q^{-\varepsilon}. \tag{4}$$

The second form follows easily from the first, by virtue of the inequality

$$L'(s, \chi) = O(\log^2 q)$$

for $1 - 1/\log q \leq s \leq 1$, which was (11) of §14. If q is large, as we may suppose, then a zero β of $L(s, \chi)$ satisfying (4) will lie in the interval just specified, and it will follow that

$$L(1, \chi) = L(1, \chi) - L(\beta, \chi) < c_1(\log^2 q)C_3(\varepsilon)q^{-\varepsilon},$$

which contradicts (1) if we there replace ε by $\frac{1}{2}\varepsilon$. This proves the second form of the theorem (assuming the first) when χ is primitive, and this suffices to prove it when χ is any real nonprincipal character.

[1] *Acta Arithmetica*, **1**, 83–86 (1935).

It follows that any real zero β of $L(s, \chi)$, for real nonprincipal χ, satisfies

(5) $$\beta \leq 1 - C_3(\varepsilon)q^{-\varepsilon},$$

and this is a much superior estimate, in principle, to any we have had hitherto. It has, however, the disadvantage of being noneffective, in the sense that it is not possible, with existing knowledge, to assign a numerical value to $C_3(\varepsilon)$ for a particular value of ε (for example, $\frac{1}{4}$).

Siegel's theorem was the culmination of a series of discoveries by several mathematicians. The problem of proving that $h(d) \to \infty$ as $d \to -\infty$, or even of proving that $h(d) \geq 2$ if $-d$ is sufficiently large, was propounded by Gauss, but no progress toward its solution was made until modern times. Hecke[2] proved that if the inequality $\beta < 1 - c_2/\log q$ holds for the real zeros of L functions formed with real primitive characters, then $h(d) > c_3|d|^{\frac{1}{2}}/\log |d|$. In particular this conclusion would follow from the generalized Riemann hypothesis.

In 1933, Deuring[3] proved the unexpected result that the *falsity* of the classical Riemann hypothesis for $\zeta(s)$ implies that $h(d) \geq 2$ if $-d$ is sufficiently large, and shortly afterward Mordell[4] proved that this assumption also implies that $h(d) \to \infty$ as $d \to -\infty$. Their work was based on a study of the behavior, as $d \to -\infty$, of

$$\sum_{Q} \sum_{x,y} Q(x, y)^{-s},$$

where Q runs through a representative set of forms of discriminant d.

In 1934, Heilbronn[5] took a further important step forward. He proved that the falsity of the generalized Riemann hypothesis implies that $h(d) \to \infty$ as $d \to -\infty$. Combined with the result of Hecke, this gave an unconditional proof that $h(d) \to \infty$, and so solved Gauss' problem.

Also in 1934, Heilbronn and Linfoot[6] proved that there are at most ten negative discriminants d for which $h(d) = 1$. As nine such d were known,

$$-3, -4, -7, -8, -11, -19, -43, -67, -163,$$

the question was whether there is a tenth such discriminant. If there were, then the L function $L(s, \chi_d)$ would have a real zero β larger

[2] See Landau, *Göttinger Nachrichten*, **1918**, 285–295. The same argument allows one to deduce the first form of Siegel's theorem from the second.

[3] *Math. Zeitschrift*, **37**, 405–415 (1933).

[4] *J. London Math. Soc.*, **9**, 289–298 (1934).

[5] *Quarterly J. of Math.*, **5**, 150–160 (1934).

[6] *Quarterly J. of Math.*, **5**, 293–301 (1934).

than $\frac{1}{2}$. In 1966, Baker[7] and Stark[8] proved independently that there is no such tenth discriminant. Baker noted that his fundamental theorem in transcendence theory provides a solution of this class number problem in view of earlier work of Gelfond and Linnik. Stark was inspired by a paper of Heegner[9] in which elliptic modular functions were used to show that there is no tenth discriminant with class number 1. It was long thought that Heegner's argument was incomplete, partly because it seemed to depend on an unproved conjecture of Weber. However, in retrospect it has now been found that Heegner's proof is essentially correct; the obscure details have been clarified by Deuring[10] and Stark[11].

Baker and Stark have found[12] all quadratic discriminants $d < 0$ for which $h(d) = 2$, but for $h(d) = 3$ the problem of finding all such d is still open. The fact that it has not been possible to find all such d, or to reduce the problem to one of computation, reflects the fact that the more powerful arguments that have been developed for this problem are of an indirect character.

We shall now prove Siegel's theorem, in the form first stated, using the simplified method given later by Estermann.[13] The basic idea is to combine the L functions of two characters. Let χ_1, χ_2 be real primitive characters to the distinct moduli q_1, q_2; as we saw in §14, $\chi_1\chi_2$ is a nonprincipal (though not necessarily primitive) character to the modulus $q_1 q_2$. Let

$$F(s) = \zeta(s)L(s, \chi_1)L(s, \chi_2)L(s, \chi_1\chi_2). \tag{6}$$

Then $F(s)$ is regular in the whole plane except for a simple pole at $s = 1$, and its residue at this pole is

$$\lambda = L(1, \chi_1)L(1, \chi_2)L(1, \chi_1\chi_2). \tag{7}$$

An essential part in the proof is played by the inequality

$$F(s) > \tfrac{1}{2} - \frac{c_4\lambda}{1 - s}(q_1 q_2)^{8(1-s)} \qquad \text{for } \tfrac{7}{8} < s < 1. \tag{8}$$

An inequality of the same general character as (8) was used by Siegel, and was deduced by him from the work of Hecke on the

[7] *Mathematika*, **13**, 204–216 (1966). See also Chapter 5 of Baker, *Transcendental Number Theory*, Cambridge University Press, 1975.

[8] *Michigan Math. J.*, **14**, 1–27 (1967).

[9] *Math. Z.*, **56**, 227–252 (1952).

[10] *Invent. Math.*, **5**, 169–179 (1968).

[11] *J. Number Theory*, **1**, 16–27 (1969); *Modular Functions of One Variable* I, Springer-Verlag, Berlin, 1973, pp. 153–174.

[12] *Ann. Math.*, **94**, 139–152, 153–173 (1971).

[13] *J. London Math. Soc.*, **23**, 275–279 (1948). Other simple proofs have been given by Chowla, *Annals of Math.* (2) **51**, 120–122 (1950) and by Goldfeld, *Proc. Nat. Acad. Sci. U.S.A.*, **71**, 1055 (1974).

functional equation of the Dedekind ζ function of an arbitrary algebraic number field. The function $F(s)$ is essentially the Dedekind ζ function of a biquadratic field. A simple proof of Siegel's inequality was given by Heilbronn,[14] but even this requires some knowledge of algebraic number theory and contains some complications of detail.

Estermann's proof of (8) is relatively simple. The multiplication of the Euler products gives

$$F(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

for $\sigma > 1$, where $a_1 = 1$ and $a_n \geq 0$ for all n. The last fact follows from

$$\log F(s) = \sum_p \sum_{m=1}^{\infty} m^{-1} p^{-ms}[1 + \chi_1(p^m)][1 + \chi_2(p^m)],$$

where the coefficients are obviously nonnegative. As in de la Vallée Poussin's argument in §4, we obtain

$$F(s) = \sum_{m=0}^{\infty} b_m(2 - s)^m$$

for $|s - 2| < 1$, where $b_0 \geq 1$ and $b_m \geq 0$ for all m.

It follows that

$$F(s) - \lambda/(s - 1) = \sum_{m=0}^{\infty} (b_m - \lambda)(2 - s)^m, \tag{9}$$

and this must be valid for $|s - 2| < 2$, since the left side is regular there.

On the circumference of the circle $|s - 2| = \frac{3}{2}$, the function $\zeta(s)$ is bounded, and the L functions satisfy

$$|L(s, \chi_1)| < c_5 q_1, \qquad |L(s, \chi_2)| < c_5 q_2, \qquad |L(s, \chi_1\chi_2)| < c_5 q_1 q_2,$$

by (14) of §12. (This inequality was proved for any nonprincipal character, whether primitive or not.) Hence

$$|F(s)| < c_6 q_1^2 q_2^2$$

on the circumference, and the same applies to $\lambda/(s - 1)$, since λ is the product of three L functions which satisfy the above inequalities. It follows from Cauchy's inequalities for the coefficients of a power series, applied to the function (9), that

$$|b_m - \lambda| < 2c_6 q_1^2 q_2^2 (\tfrac{2}{3})^m. \tag{10}$$

[14] *Quarterly J. of Math.*, 9, 194–195 (1938).

For $\frac{7}{8} \le s < 1$, we have

$$\sum_{m=M}^{\infty} |b_m - \lambda|(2 - s)^m \le \sum_{m=M}^{\infty} 2c_6 q_1^2 q_2^2 [\tfrac{2}{3}(2 - s)]^m$$
$$\le 2c_6 q_1^2 q_2^2 \sum_{m=M}^{\infty} (\tfrac{3}{4})^m$$
$$< c_7 q_1^2 q_2^2 (\tfrac{3}{4})^M < c_7 q_1^2 q_2^2 e^{-M/4}.$$

Hence, in this interval,

$$F(s) - \lambda/(s - 1) \ge 1 - \lambda \sum_{m=0}^{M-1} (2 - s)^m - c_7 q_1^2 q_2^2 e^{-M/4}$$
$$= 1 - \lambda \frac{(2 - s)^M - 1}{1 - s} - c_7 q_1^2 q_2^2 e^{-M/4}.$$

We choose M to satisfy

$$\tfrac{1}{2}e^{-1/4} \le c_7 q_1^2 q_2^2 e^{-M/4} < \tfrac{1}{2}$$

and obtain

$$F(s) > \tfrac{1}{2} - \frac{\lambda}{1 - s}(2 - s)^M.$$

Since

$$\tfrac{1}{4}M \le 2 \log q_1 q_2 + c_8,$$

so that

$$M \le 8 \log q_1 q_2 + c_9,$$

we have

$$(2 - s)^M = \exp[M \log(1 + 1 - s)] < \exp[M(1 - s)] < c_{10}(q_1 q_2)^{8(1-s)}.$$

This proves (8).

To deduce Siegel's theorem, we distinguish (following Estermann) two cases, the distinction depending on the given positive number ε. If there is a real primitive character for which $L(s, \chi)$ has a real zero between $1 - \frac{1}{16}\varepsilon$ and 1, we choose χ_1 to be such a character and β_1 to be the zero in question. Then $F(\beta_1) = 0$, independently of what χ_2 may be. If there is no such character, we choose χ_1 to be any real primitive character and β_1 to be any number satisfying $1 - \frac{1}{16}\varepsilon < \beta_1 < 1$. Then $F(\beta_1) < 0$, independently of what χ_2 may be, for $\zeta(s) < 0$ when $0 < s < 1$ by (7) of §4, and the three L functions in (6) are positive when $s = 1$ and do not vanish for $\beta_1 \le s \le 1$. Thus in either case $F(\beta_1) \le 0$, and the inequality (8) gives

$$c_4 \lambda > \tfrac{1}{2}(1 - \beta_1)(q_1 q_2)^{-8(1-\beta_1)}.$$

From now on we keep χ_1 and β_1 fixed. Let χ_2 be any real primitive character to a modulus $q_2 > q_1$. It follows from (13) of §14 that

$$\lambda < (c_{11} \log q_1)L(1, \chi_2)(c_{11} \log q_1 q_2).$$

Hence

$$L(1, \chi_2) > Cq_2^{-8(1-\beta_1)}(\log q_2)^{-1},$$

where C depends only on χ_1, and therefore only on ε. Since $8(1 - \beta_1) < \frac{1}{2}\varepsilon$, the last inequality implies (1) if q_2 is sufficiently large (as we may suppose). This establishes Siegel's theorem.

22

THE PRIME NUMBER THEOREM FOR ARITHMETIC PROGRESSIONS (II)

By appealing to Siegel's theorem we can obtain a better approximation to $\psi(x; q, a)$ than was possible in §20.

If we suppose that

$$(1) \qquad q \le \exp[C(\log x)^{\frac{1}{2}}]$$

for some positive constant C, then (8) of §20 tells us that (for $\chi \neq \chi_0$)

$$\psi(x, \chi) = -\frac{x^{\beta_1}}{\beta_1} + O\{x \exp[-C'(\log x)^{\frac{1}{2}}]\},$$

where C' is a positive constant depending only on C. Here the term in β_1 occurs for at most one real character (mod q). Siegel's theorem states that for any $\varepsilon > 0$ there exists $C_1(\varepsilon)$ such that

$$\beta_1 < 1 - C_1(\varepsilon) q^{-\varepsilon}.$$

Hence

$$x^{\beta_1} < x \exp[-C_1(\varepsilon)(\log x) q^{-\varepsilon}].$$

In order that this expression may be small compared with x, we must impose a more severe restriction on q than that expressed by (1). Suppose that

$$(2) \qquad q \le (\log x)^N,$$

for some positive constant N. Then, on taking $\varepsilon = (2N)^{-1}$, we get $q^\varepsilon \le (\log x)^{\frac{1}{2}}$, and

$$x^{\beta_1} < x \exp[-C_2(N)(\log x)^{\frac{1}{2}}].$$

Thus, subject to (2), we have

$$(3) \qquad |\psi(x, \chi)| \ll x \exp[-C_3(N)(\log x)^{\frac{1}{2}}],$$

for any nonprincipal χ (mod q).

Substituting in (3) of §20, we obtain the following result for $\psi(x; q, a)$, which represents the best form so far known of the prime number theorem for arithmetic progressions.[1]

Let N be any positive constant. Then there exists a positive number $C_3(N)$, depending only on N, such that if q satisfies (2) *then*

$$\psi(x; q, a) = \frac{x}{\phi(q)} + O\{x \exp[-C_3(N)(\log x)^{\frac{1}{2}}]\} \tag{4}$$

uniformly in q.

The various results for $\psi(x; q, a)$, which have been found in §20 and here, have analogs for $\pi(x; q, a)$, the number of primes up to x that are congruent to $a \pmod q$. These are derived by partial summation, as in §18. The main term is now $(\operatorname{li} x)/\phi(q)$ in place of $x/\phi(q)$, and the error terms are all reduced by a factor $\log x$. But the latter change is of no significance except for the analog of (14) of §20, which was based on the assumption of the generalized Riemann hypothesis.

As we have seen, the main difficulty in approximating to $\psi(x; q, a)$ arises from the term containing β_1. But if this term is retained, so that one is prepared to accept a result of the form

$$\psi(x; q, a) = \frac{x}{\phi(q)} - \frac{\bar{\chi}_1(a)}{\phi(q)} \frac{x^{\beta_1}}{\beta_1} + O(...), \tag{5}$$

where χ_1 is the possible real character with the exceptional zero β_1, further progress is possible. The error term then comes essentially from

$$\frac{1}{\phi(q)} \sum_{\chi} \sum_{\rho} |x^{\rho}/\rho|,$$

and here it is not essential to have a good estimate for the real part of each ρ, provided one can handle the above average over the $\phi(q)$ characters. Results in this direction can be based on estimates for

$$\sum_{\chi} N(\alpha, T, \chi), \tag{6}$$

where $N(\alpha, T, \chi)$ denotes the number of zeros of $L(s, \chi)$ in the rectangle

$$\alpha \le \sigma < 1, \qquad |t| < T.$$

Such estimates were obtained[2] by Rodosskii and Tatuzawa, building upon work of Linnik, and as a consequence they were able to obtain

[1] This application of Siegel's theorem to primes in arithmetic progressions was made by Walfisz, *Math. Zeitschrift*, **40**, 592–607 (1936).

[2] For an account of their work, see Prachar, Chap. 9.

an improved error term in (5), or alternatively the same error term for a longer range of q.

The value of a result of the type (5) is mainly in connection with the distribution of primes in a relatively short segment of an arithmetic progression. When such a formula is applied with two values of x that are not far apart, and the results subtracted, the terms containing β_1 largely cancel.

The methods for estimating the sum (6) are based to a considerable extent on earlier work[3] by a large number of mathematicians on the estimation of $N(\alpha, T)$, the number of zeros of $\zeta(s)$ in the rectangle $\alpha \le \sigma < 1$, $0 < t < T$.

From now on, we shall be concerned primarily with the proof of an estimate for

$$\psi(x; q, a) - x/\phi(q),$$

not for an individual value of q but on the average over q up to a certain bound. Such results are obtained by the "large sieve" method, which we discuss in §27.

The first result of this general nature was given by Rényi. He proved[4] that, for primes $q \le (\frac{1}{12}N)^{\frac{1}{3}}$, the inequality

$$|\pi(x; q, a) - (\operatorname{li} x)/\phi(q)| \ll \frac{N}{q^{\frac{5}{4}} \log N}$$

holds, apart from certain possible exceptional pairs q, a; the number of exceptional q is $\ll N^{\frac{1}{4}} \log N$, and the number of exceptional a for each q is $\ll q^{\frac{3}{4}}$.

In §28 we shall prove the following simple and far-reaching result of Bombieri: For any positive constant A, there exists a positive constant B such that

$$\sum_{q \le X} \max_{(a,q)=1} \max_{y \le x} |\psi(y; q, a) - y/\phi(q)| \ll x(\log x)^{-A}$$

if

$$X = x^{\frac{1}{2}}(\log x)^{-B}.$$

The form of the inequality, with two maxima, may at first sight seem complicated, but it is one that is very convenient for applications.

[3] See Titchmarsh, Chap. 9.

[4] *Compositio Mathematica*, **8**, 68–75 (1950).

23

THE PÓLYA–VINOGRADOV INEQUALITY

Suppose that χ is a nonprincipal character (mod q). Since $\sum_{n=1}^{q} \chi(n) = 0$, it is clear that $\sum_{n=M+1}^{M+N} \chi(n) \ll q$ for any M and N. However, a sharper bound is needed to describe the distribution of power residues within the interval $1 \le n \le q$. In 1918 Pólya[1] and Vinogradov[2] proved independently that

$$\sum_{n=M+1}^{M+N} \chi(n) \ll q^{\frac{1}{2}} \log q \tag{1}$$

for nonprincipal characters χ (mod q). By taking $\chi(n) = (n|p)$, we deduce that the interval $M+1 \le n \le M+N$ contains $\frac{1}{2}N + O(p^{\frac{1}{2}} \log p)$ quadratic residues (mod p). The Pólya–Vinogradov inequality will be used in our arguments of §28.

Pólya considered the sum $\sum_{n \le xq} \chi(n)$ as a function with period 1, and determined its Fourier expansion. The Fourier expansion is not absolutely convergent, and so does not immediately provide a proof of (1), but Pólya also derived a truncated expansion which suffices. Pólya's analysis is fundamental to more detailed investigations, but for our purposes an elementary argument of Schur[3] suffices.

We first prove that

$$\left| \sum_{n=M+1}^{M+N} \chi(n) \right| < q^{\frac{1}{2}} \log q \tag{2}$$

for primitive characters χ (mod q), $q > 1$. In §9 we saw that for such χ and any n,

$$\chi(n) = \frac{1}{\tau(\bar{\chi})} \sum_{a=1}^{q} \bar{\chi}(a) e\left(\frac{an}{q}\right).$$

[1] *Göttinger Nachrichten*, **1918**, 21–29.
[2] *Perm. Univ. Fiz.-mat. ob-vo Zh.*, **1**, 18–28, 94–98 (1918).
[3] *Göttinger Nachrichten*, **1918**, 30–36.

Hence the sum in question is

$$\frac{1}{\tau(\bar{\chi})}\sum_{a=1}^{q}\bar{\chi}(a)\sum_{n=M+1}^{M+N}e\left(\frac{an}{q}\right).$$

Here $|\tau(\bar{\chi})| = q^{1/2}$, and the inner sum is a geometric series with sum

$$= e\left(\frac{(M+\frac{1}{2}N+\frac{1}{2})a}{q}\right)\frac{\sin \pi Na/q}{\sin \pi a/q}.$$

Consequently

$$\left|\sum_{n=M+1}^{M+N}\chi(n)\right| \le q^{-\frac{1}{2}}\sum_{a=1}^{q-1}\frac{1}{|\sin \pi a/q|}.$$

For convex functions $f(\alpha)$,

$$f(\alpha) \le \frac{1}{\delta}\int_{\alpha-\frac{1}{2}\delta}^{\alpha+\frac{1}{2}\delta} f(\beta)\,d\beta.$$

Taking $f(\alpha) = (\sin \pi\alpha)^{-1}$, $\delta = 1/q$, we see that the sum above is

$$\le q\int_{\frac{1}{2}q}^{1-(\frac{1}{2}q)}(\sin \pi\beta)^{-1}\,d\beta = 2q\int_{\frac{1}{2}q}^{\frac{1}{2}}(\sin \pi\beta)^{-1}\,d\beta.$$

Now $\sin \pi\beta > 2\beta$ for $0 < \beta < \frac{1}{2}$, so that the above is

$$< 2q\int_{\frac{1}{2}q}^{\frac{1}{2}}\frac{d\beta}{2\beta} = q\log q.$$

Hence we have (2) for primitive χ.

Suppose now that χ is a nonprincipal character (mod q), induced by the primitive character $\chi_1(\text{mod } q_1)$. Then $q_1 | q$, and we write $q = q_1 r$. Hence

$$\sum_{n=M+1}^{M+N}\chi(n) = \sum_{\substack{n=M+1\\(n,r)=1}}^{M+N}\chi_1(n).$$

Now $\sum_{d|n}\mu(d) = 1$ or 0 according as $n = 1$ or $n > 1$, so that the above is

$$\sum_{n=M+1}^{M+N}\chi_1(n)\sum_{\substack{d|n\\d|r}}\mu(d) = \sum_{d|r}\mu(d)\sum_{\substack{n=M+1\\d|n}}^{M+N}\chi_1(n)$$

$$= \sum_{d|r}\mu(d)\chi_1(d)\sum_{(M+1)/d\le m\le(M+N)/d}\chi_1(m).$$

In view of (2), the inner sum has modulus $< q_1^{1/2}\log q_1$, so that

$$\left|\sum_{n=M+1}^{M+N}\chi(n)\right| < q_1^{\frac{1}{2}}(\log q_1)\sum_{d|r}|\mu(d)| = 2^{\omega(r)}q_1^{\frac{1}{2}}\log q_1.$$

But $2^{\omega(r)} \le d(r) \ll r^{\varepsilon}$ for any $\varepsilon > 0$, and in particular for $\varepsilon = \frac{1}{2}$, which gives (1). In fact we can obtain a good numerical constant by noting that

$$d(r) = \sum_{d \mid r} 1 \le 2 \sum_{\substack{d \mid r \\ d \le \sqrt{r}}} 1 \le 2\sqrt{r}.$$

The inequality (1) is close to best possible, for Schur also proved that

$$\max_N \left| \sum_{n \le N} \chi(n) \right| > \frac{1}{2\pi} \sqrt{q}$$

for all primitive χ (mod q). In 1932 Paley[4] showed that

$$\max_N \left| \sum_{n \le N} \left(\frac{d}{n} \right) \right| > \tfrac{1}{7}\sqrt{d} \log \log d$$

for infinitely many quadratic discriminants $d > 0$. In the opposite direction Montgomery and Vaughan[5] have shown recently that, assuming the generalized Riemann hypothesis,

$$\sum_{n=M+1}^{M+N} \chi(n) \ll \sqrt{q} \log \log q$$

for all $\chi \ne \chi_0 (\text{mod } q)$. Although (1) is close to being best possible, for many purposes it is useful to have an estimate which is sharper when N is small compared with q; Burgess[6] has made some progress in this direction.

[4] *J. London Math. Soc.*, **7**, 28–32 (1932).
[5] *Invent. Math.*, **43**, 69–82 (1977).
[6] *Proc. London Math. Soc.*, (3) **13**, 524–536 (1963).

24

FURTHER PRIME NUMBER SUMS

When f is monotonic we can use the prime number theorem and partial summation to estimate $\sum_{p \le N} f(p)$. For certain multiplicative functions, namely those of the form $f(n) = \chi(n)n^{-s}$, we can estimate $\sum_{p \le N} f(p)$ by using the zero-free region of $L(s, \chi)$. In 1937 Vinogradov introduced a method for estimating sums $\sum_{p \le N} f(p)$ in which f is oscillatory but not multiplicative.[1] His starting point was a simple sieve idea. Let $P = \prod_{p \le N^{\frac{1}{2}}} p$. For n in the range $1 \le n \le N$ the sieve of Eratosthenes asserts that $(n, P) = 1$ if and only if $n = 1$ or n is a prime number in the interval $N^{\frac{1}{2}} < n \le N$. Hence

$$f(1) + \sum_{N^{\frac{1}{2}} < p \le N} f(p) = \sum_{\substack{n \le N \\ (n, P) = 1}} f(n) = \sum_{\substack{t | P \\ t \le N}} \mu(t) \sum_{r \le N/t} f(rt).$$

Thus we are led to bound sums of the kind $\sum_{r \le N/t} f(rt)$. We need to show that these sums are small. However, we cannot hope to get much cancellation when t is nearly as large as N, for then the sum contains few terms. Therefore Vinogradov rearranged the terms arising from $t | P$, $\delta N \le t \le N$, but this entailed great complications. Recently Vaughan[2] found a new version of Vinogradov's method in which the details are much simpler.

Following Vaughan, we let

$$F(s) = \sum_{m \le U} \Lambda(m) m^{-s}, \qquad G(s) = \sum_{d \le V} \mu(d)\, d^{-s},$$

and we note the identity

$$(1) \qquad -\frac{\zeta'(s)}{\zeta(s)} = F(s) - \zeta(s)F(s)G(s) - \zeta'(s)G(s) + \left(-\frac{\zeta'(s)}{\zeta(s)} - F(s)\right) \cdot (1 - \zeta(s)G(s)),$$

[1] See Chapter IX of Vinogradov, *The Method of Trigonometrical Sums in the Theory of Numbers*, Interscience, London, 1954.

[2] *C.R. Acad. Sci. Paris*, Sér A, **285**, 981–983 (1977).

valid for $\sigma > 1$. Calculating the Dirichlet series coefficients of the four functions on the right-hand side, we see that

$$\Lambda(n) = a_1(n) + a_2(n) + a_3(n) + a_4(n),$$

where

$$a_1(n) = \begin{cases} \Lambda(n) & \text{if } n \le U, \\ 0 & \text{if } n > U; \end{cases}$$

$$a_2(n) = -\sum_{\substack{mdr=n \\ m \le U \\ d \le V}} \Lambda(m)\mu(d);$$

$$a_3(n) = \sum_{\substack{hd=n \\ d \le V}} \mu(d) \log h;$$

and

$$a_4(n) = -\sum_{\substack{mk=n \\ m>U \\ k>1}} \Lambda(m)\Bigg(\sum_{\substack{d \mid k \\ d \le V}} \mu(d)\Bigg).$$

We multiply throughout by $f(n)$ and sum; then

$$\sum_{n \le N} f(n)\Lambda(n) = S_1 + S_2 + S_3 + S_4,$$

where

$$S_i = \sum_{n \le N} f(n)a_i(n).$$

In applications we shall bound S_1 trivially; the remaining S_i are treated individually.

We write S_2 in the form

$$S_2 = -\sum_{t \le UV}\Bigg(\sum_{\substack{md=t \\ m \le U \\ d \le V}} \mu(d)\Lambda(m)\Bigg)\sum_{r \le N/t} f(rt).$$

Again we have a linear combination of the sums $\sum_{r \le N/t} f(rt)$, but now we can control the range of t by ensuring that UV is substantially smaller than N. As $\sum_{m \mid t} \Lambda(m) = \log t \le \log UV$, we see that

$$S_2 \ll (\log UV) \sum_{t \le UV}\Bigg|\sum_{r \le N/t} f(rt)\Bigg|. \tag{2}$$

The sum S_3 is of the same form, since

$$S_3 = \sum_{d \le V} \mu(d) \sum_{h \le N/d} f(dh) \log h = \sum_{d \le V} \mu(d) \sum_{h \le N/d} f(dh) \int_1^h \frac{dw}{w}$$

$$= \int_1^N \sum_{d \le V} \mu(d) \sum_{w \le h \le N/d} f(dh) \frac{dw}{w} \tag{3}$$

$$\ll (\log N) \sum_{d \le V} \max_w \left| \sum_{w \le h \le N/d} f(dh) \right|.$$

The sum S_4 has a more complicated shape. We note that

$$\sum_{\substack{d|k \\ d \le V}} \mu(d) = 0$$

for $1 < k \le V$, so that

$$S_4 = \sum_{U < m \le N/V} \Lambda(m) \sum_{V < k \le N/m} \left(\sum_{\substack{d|k \\ d \le V}} \mu(d) \right) f(mk).$$

Suppose that $\Delta = \Delta(f, M, N, V)$ is such that

$$\left| \sum_{M < m \le 2M} b_m \sum_{V < k \le N/m} c_k f(mk) \right| \le \Delta \left(\sum_M^{2M} |b_m|^2 \right)^{\frac{1}{2}} \left(\sum_{k \le N/M} |c_k|^2 \right)^{\frac{1}{2}} \tag{4}$$

for any complex numbers b_m, c_k; such bilinear form inequalities are familiar, and we have means of estimating Δ. Thus

$$S_4 \ll (\log N) \max_{U \le M \le N/V} \Delta \left(\sum_M^{2M} \Lambda(m)^2 \right)^{\frac{1}{2}} \left(\sum_{k \le N/M} d(k)^2 \right)^{\frac{1}{2}}.$$

Here the sum over m is estimated by noting that

$$\sum_{m \le z} \Lambda(m)^2 \le (\log z) \sum_{m \le z} \Lambda(m) \ll z \log z.$$

As for the sum over k, we observe that $d(k)^2 = \sum_{d|k} f(d)$, where $f(d)$ is the multiplicative function for which $f(p^a) = 2a + 1$. Thus

$$\sum_{k \le z} d(k)^2 = \sum_{k \le z} \sum_{d|k} f(d) = \sum_{d \le z} f(d)[z/d]$$

$$\le z \sum_{d \le z} f(d)/d \le z \prod_{p \le z} (1 + f(p)/p + f(p^2)/p^2 + \cdots)$$

$$\le z \prod_{p \le z} \left(1 - \frac{1}{p}\right)^{-3} \ll z(\log 2z)^3.$$

Combining these estimates, we see that

$$S_4 \ll N^{\frac{1}{2}}(\log N)^3 \max_{U \le M \le N/V} \Delta. \tag{5}$$

To be more specific we now suppose that $|f(n)| \le 1$ for all n. Then the estimate $\sum_{n \le N} f(n)\Lambda(n) \ll N$ is trivial, and we seek a sharper estimate. Clearly $S_1 \ll U$. From (2) we obtain the trivial estimate $S_2 \ll N(\log UV)^2$; hence we do not require much cancellation in the sums $\sum_{t \le N/r} f(rt)$ to show that $S_2 = o(N)$. Similar remarks apply to S_3. For S_4 we obtain a trivial bound for Δ by applying Cauchy's inequality:

$$\sum_{M < m \le 2M} b_m \sum_{V < k \le N/M} c_k \ll \left(M \cdot \frac{N}{M}\right)^{\frac{1}{2}} \left(\sum_M^{2M} |b_m|^2\right)^{\frac{1}{2}} \cdot \left(\sum_{k \le N/M} |c_k|^2\right)^{\frac{1}{2}}.$$

Hence the estimate $\Delta \ll N^{\frac{1}{2}}$ is trivial, which in (5) gives $S_4 \ll N(\log N)^3$; thus we need only a slightly sharper bound for Δ. Note however that no such improvement is possible if f is totally multiplicative and unimodular, since we may take $b_m = \overline{f(m)}$, $c_k = \overline{f(k)}$. For this reason the principal applications of the method involve functions f which are not multiplicative.

For most f we are not able to determine the least Δ for which (4) holds. However, the following approach is very useful: By Cauchy's inequality, the left-hand side of (4) is

$$\le \left(\sum_M^{2M} |b_m|^2\right)^{\frac{1}{2}} \left(\sum_M^{2M} \left| \sum_{V < k \le N/m} c_k f(mk) \right|^2\right)^{\frac{1}{2}}.$$

Here the second sum over m is

$$\sum_{V < j \le N/M} c_j \sum_{V < k \le N/M} \bar{c}_k \sum_{\substack{M < m \le 2M \\ m \le N/j \\ m \le N/k}} f(mj)\overline{f(mk)}.$$

We note that $|c_j \bar{c}_k| \le \frac{1}{2}|c_j|^2 + \frac{1}{2}|c_k|^2$; hence the above is

$$\ll \sum_{V < j \le N/M} |c_j|^2 \sum_{V < k \le N/M} \left| \sum_{\substack{M \le m \le 2M \\ m \le N/j \\ m \le N/k}} f(mj)\overline{f(mk)} \right|.$$

Thus

$$\Delta \ll \left(\max_{V < j \le N/M} \sum_{V < k \le N/M} \left| \sum_{\substack{M < m \le 2M \\ m \le N/j \\ m \le N/k}} f(mj)\overline{f(mk)} \right| \right)^{\frac{1}{2}}.$$

This bound is largest when $f \equiv 1$, and then we obtain again the trivial bound $\Delta \ll N^{\frac{1}{2}}$. If f is totally multiplicative and unimodular

the bound is unchanged, but otherwise we may expect some cancellation in the inner sum, and hence a nontrivial bound for Δ.

Combining our estimates, we see that if $|f(n)| \le 1$ for all n, $U \ge 2$, $V \ge 2$, $UV \le N$, then

$$(6) \qquad \sum_{n \le N} f(n)\Lambda(n) \ll U + (\log N) \sum_{t \le UV} \max_{w} \left| \sum_{w \le r \le N/t} f(rt) \right|$$

$$+ N^{\frac{1}{2}}(\log N)^3 \max_{U \le M \le N/V} \max_{V \le j \le N/M} \left(\sum_{V < k \le N/M} \left| \sum_{\substack{M < m \le 2M \\ m \le N/k \\ m \le N/j}} f(mj)\overline{f(mk)} \right| \right)^{\frac{1}{2}}.$$

In conclusion we remark that in some situations sharper estimates can be obtained by treating S_2 more carefully. Write

$$S_2 = \sum_{t \le UV} = \sum_{t \le U} + \sum_{U < t \le UV} = S_2' + S_2''.$$

Then treat S_2' as we did S_2, and S_2'' as we did S_4.

25

AN EXPONENTIAL SUM FORMED WITH PRIMES

Vinogradov first used his method to estimate the important sum

$$S(\alpha) = \sum_{n \leq N} \Lambda(n) e(n\alpha).$$

We now use our general estimates of the previous section to bound $S(\alpha)$. We find that our results depend on rational approximations to α: If

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{q^2}, \quad (a, q) = 1, \tag{1}$$

then

$$S(\alpha) \ll (Nq^{-\frac{1}{2}} + N^{\frac{4}{5}} + N^{\frac{1}{2}}q^{\frac{1}{2}})(\log N)^4. \tag{2}$$

To prove this we first note that

$$\sum_{N_1}^{N_2} e(n\beta) = \frac{e((N_2 + 1)\beta) - e(N_1\beta)}{e(\beta) - 1} \ll \min\left(N_2 - N_1, \frac{1}{\|\beta\|}\right).$$

Here $\|\beta\|$ denotes the distance from β to the nearest integer. Hence

$$\sum_{t \leq T} \max_{w} \left| \sum_{w \leq r \leq N/t} e(rt\alpha) \right| \ll \sum_{t \leq T} \min\left(\frac{N}{t}, \frac{1}{\|t\alpha\|}\right).$$

We assume for the moment that this latter expression is

$$\ll \left(\frac{N}{q} + T + q\right) \log 2qT \tag{3}$$

for α satisfying (1). Then the upper bound (6) of §24 gives

$$S(\alpha) \ll U + \left(\frac{N}{q} + UV + q\right)(\log 2qN)^2$$

$$+ N^{1/2}(\log N)^3 \max_{U \leq M \leq N/V} \max_{V < j \leq N/M} \left(\sum_{V < k \leq N/M} \min\left(M, \frac{1}{\|(k - j)\alpha\|}\right)\right)^{\frac{1}{2}}.$$

This last term is

$$\ll N^{\frac{1}{2}}(\log N)^3 \max_{U \le M \le N/V} \left(M + \sum_{1 \le m \le N/M} \min\left(\frac{N}{m}, \frac{1}{\|m\alpha\|}\right)\right)^{\frac{1}{2}},$$

and by (3) again this is

$$\ll N^{\frac{1}{2}}(\log N)^3 \max_{U \le M \le N/V} \left(M + \frac{N}{M} + \frac{N}{q} + q\right)^{\frac{1}{2}} (\log qN)^{\frac{1}{2}}$$

$$\ll (NV^{-\frac{1}{2}} + NU^{-\frac{1}{2}} + Nq^{-\frac{1}{2}} + N^{\frac{1}{2}}q^{\frac{1}{2}})(\log qN)^4.$$

Hence altogether we have

$$S(\alpha) \ll (UV + q + NU^{-\frac{1}{2}} + NV^{-\frac{1}{2}} + Nq^{-\frac{1}{2}} + N^{\frac{1}{2}}q^{\frac{1}{2}})(\log qN)^4.$$

The estimate (2) is trivial if $q > N$, so we may assume that $q \le N$. Then we obtain (2) by taking $U = V = N^{\frac{2}{5}}$.

It remains to establish the estimate (3). Write $t = hq + r$ with $1 \le r \le q$, and put $\beta = \alpha - a/q$. Then

$$\sum_{t \le T} \min\left(\frac{N}{t}, \frac{1}{\|t\alpha\|}\right) \ll \sum_{0 \le h \le T/q} \sum_{r=1}^{q} \min\left(\frac{N}{hq + r}, \|ra/q + hq\beta + r\beta\|^{-1}\right).$$

We consider first those terms for which $h = 0$, $1 \le r \le \frac{1}{2}q$. For such r we have $|r\beta| \le 1/2q$, so that the contribution of these terms is

$$\ll \sum_{1 \le r \le q/2} \frac{1}{\left\|\dfrac{ra}{q}\right\| - \dfrac{1}{2q}} \ll q \log q.$$

For all remaining terms we have $hq + r \gg (h + 1)q$. Let h be given, and let I be an interval in $[0, 1]$ of length $1/q$. There are at most 4 values of r, $1 \le r \le q$, for which

$$\frac{ra}{q} + hq\beta + r\beta \in I \pmod 1.$$

Hence

$$\sum_{0 \le h \le T/q} \sum_{r=1}^{q} \min\left(\frac{N}{(h+1)q}, \left\|\frac{ra}{q} + hq\beta + r\beta\right\|^{-1}\right)$$

$$\ll \sum_{0 \le h \le T/q} \left(\frac{N}{(h+1)q} + q \log 2q\right)$$

$$\ll \left(\frac{N}{q} + T + q\right) \log 2qT,$$

and the proof is complete.

One may note that our estimate (3) is sharp, even in the special case $\alpha = a/q$, but that if the hypothesis (1) is weakened then the bound (3) must be correspondingly weakened.

26

SUMS OF THREE PRIMES

Hardy and Littlewood[1] showed, assuming the generalized Riemann hypothesis, that every sufficiently large odd number is a sum of three primes. In their argument, the hypothesis was required to provide estimates corresponding to our estimates of $S(\alpha)$ in §25. In 1937 Vinogradov[2] used his new estimates to treat sums of three primes unconditionally. Instead of considering the number of representations of n as a sum of three primes, we deal with the related quantity

$$r(n) = \sum \Lambda(k_1)\Lambda(k_2)\Lambda(k_3),$$

where the sum is extended over all triples k_1, k_2, k_3 of numbers for which $k_1 + k_2 + k_3 = n$. Thus $r(n)$ is a weighted counting of the number of representations of n as a sum of three prime powers. In additive questions it is appropriate to use a power-series generating function or exponential sum. Taking

$$S(\alpha) = \sum_{k \le N} \Lambda(k)e(k\alpha),$$

we see that

$$S(\alpha)^3 = \sum_{n} r'(n)e(n\alpha),$$

where $r'(n)$ is defined in the same way as $r(n)$ but with the further restriction that the k_i do not exceed N. Hence $r'(n) = r(n)$ for $n \le N$. As $S(\alpha)^3$ is a trigonometric polynomial, we can calculate $r(N)$ by the Fourier coefficient formula

$$(1) \qquad r(N) = \int_0^1 S(\alpha)^3 e(-N\alpha)\, d\alpha.$$

[1] *Acta Math.*, **44**, 1–70 (1922).
[2] *Mat. Sb.*, N.S. **2** (O.S. **44**), 179–195 (1937).

We shall find that the integrand is large when α is near a rational number with small denominator; by estimating the contributions made by these peaks, we prove the following:

THEOREM (Vinogradov). *For any fixed $A > 0$,*

$$r(N) = \tfrac{1}{2}\mathfrak{S}(N)N^2 + O(N^2(\log N)^{-A}),$$

where

$$\mathfrak{S}(N) = \left(\prod_{p \mid N}\left(1 - \frac{1}{(p-1)^2}\right)\right)\left(\prod_{p \nmid N}\left(1 + \frac{1}{(p-1)^3}\right)\right).$$

The above is of little use when N is even, for then $\mathfrak{S}(N) = 0$, at least one of the k_i is a power of two, and hence $r(N) \ll N(\log N)^4$. However, if N is odd, then $\mathfrak{S}(N) \asymp 1$, and hence $r(N) \gg N^2$ for all large odd N. The contribution made to $r(N)$ by proper prime powers is easily seen to be $\ll N^{\frac{3}{2}}(\log N)^2$; hence all large odd N can be written as a sum of three primes in $\gg N^2(\log N)^{-3}$ ways.

We now divide the range of integration in (1) into subintervals for detailed treatment. Let $P = (\log N)^B$, $Q = N(\log N)^{-B}$, where B will be chosen later in terms of A. For $q \le P$, $1 \le a \le q$, $(a, q) = 1$, let $\mathfrak{M}(q, a)$ denote the interval $|\alpha - a/q| \le 1/Q$. Here we are considering the real numbers modulo 1, so that $\mathfrak{M}(1, 1)$ can be thought of as the interval $|\alpha| \le 1/Q$. Let $\mathfrak{M}$ denote the union of these "major arcs." We note that two major arcs $\mathfrak{M}(q, a)$ and $\mathfrak{M}(q', a')$ are disjoint if $a/q \ne a'/q'$, since

$$\left|\frac{a}{q} - \frac{a'}{q'}\right| \ge \frac{1}{qq'} \ge \frac{1}{P^2} > \frac{2}{Q}.$$

We let $\mathfrak{m}$ (standing for "minor arcs") denote the complement in $[0, 1]$ of $\mathfrak{M}$.

We now estimate the contribution of the major arcs to the integral (1). To this end we first determine the size of $S(\alpha)$ for $\alpha \in \mathfrak{M}(q, a)$. We easily see that

$$\frac{1}{\phi(q)}\sum_{\chi}\tau(\bar{\chi})\chi(h) = \begin{cases} e(h/q) & \text{if } (h, q) = 1, \\ 0 & \text{if } (h, q) > 1. \end{cases}$$

Hence if $(a, q) = 1$, $\alpha = a/q + \beta$, then

$$\sum_{\substack{k \le N \\ (k, q) = 1}} \Lambda(k)e(k\alpha) = \frac{1}{\phi(q)}\sum_{\chi}\tau(\bar{\chi})\chi(a)\sum_{k \le N}\chi(k)\Lambda(k)e(k\beta),$$

so that

$$(2)\quad S(\alpha) = \frac{1}{\phi(q)} \sum_{\chi} \tau(\bar{\chi})\chi(a) \sum_{k \le N} \chi(k)\Lambda(k)e(k\beta) + O((\log N)^2).$$

It is easy to verify that the inner sum here is

$$(3)\qquad = e(N\beta)\psi(N, \chi) - 2\pi i\beta \int_1^N e(u\beta)\psi(u, \chi)\, du.$$

If $\chi \neq \chi_0$, then by estimate (3) of §22, the above is

$$\ll (1 + |\beta| N)N \exp(-c\sqrt{\log N}).$$

To treat χ_0, we let $\psi(u, \chi_0) = [u] + R(u)$, and we put $T(\beta) = \sum_{k \le N} e(k\beta)$. Again it is easily seen that

$$T(\beta) = e(N\beta)N - 2\pi i\beta \int_1^N e(n\beta)[u]\, du.$$

By subtracting this from (3) we find that

$$\begin{aligned}\sum_{k \le N} \chi_0(k)\Lambda(k)e(k\beta) &= T(\beta) + e(N\beta)R(N) - 2\pi i\beta \int_1^N e(u\beta)R(u)\, du \\ &= T(\beta) + O((1 + |\beta| N)N \exp(-c\sqrt{\log N})).\end{aligned}$$

In §9 we saw that $\tau(\chi_0) = \mu(q)$ and that $|\tau(\chi)| \le q^{\frac{1}{2}}$ for any $\chi(\text{mod } q)$. On combining these estimates in (2) we conclude that

$$S(\alpha) = \frac{\mu(q)}{\phi(q)} T(\beta) + O((1 + |\beta| N)q^{\frac{1}{2}}N \exp(-c\sqrt{\log N})).$$

But $q \le P$ and $|\beta| \le 1/Q$ for $\alpha \in \mathfrak{M}(q, a)$, so that

$$S(\alpha) = \frac{\mu(q)}{\phi(q)} T(\beta) + O(N \exp(-c_1\sqrt{\log N}))$$

for $\alpha \in \mathfrak{M}(q, a)$. Consequently

$$S(\alpha)^3 = \frac{\mu(q)}{\phi(q)^3} T(\beta)^3 + O(N^3 \exp(-c_1\sqrt{\log N})),$$

and hence the contribution of $\mathfrak{M}(q, a)$ to the integral (1) is

$$\frac{\mu(q)}{\phi(q)^3} e\left(-\frac{aN}{q}\right) \int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\, d\beta + O(N^2 \exp(-c_2\sqrt{\log N})).$$

Summing over the various major arcs, we see that

$$(4)\quad \int_{\mathfrak{M}} S(\alpha)^3 e(-N\alpha)\,d\alpha = \sum_{q \le P} \frac{\mu(q)}{\phi(q)^3} c_q(N) \int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\,d\beta + O(N^2 \exp(-c_3\sqrt{\log N})),$$

where $c_q(n)$ is Ramanujan's sum,

$$c_q(n) = \sum_{\substack{a=1 \\ (a,q)=1}}^{q} e\left(\frac{an}{q}\right).$$

We now estimate the integral and the sum occurring on the right-hand side of (4). The sum $T(\beta)$ is a geometric series with sum

$$\frac{e((N+1)\beta) - 1}{e(\beta) - 1} \ll \min(N, \|\beta\|^{-1}).$$

Hence

$$\int_{1/Q}^{1-1/Q} |T(\beta)|^3\,d\beta \ll Q^2 \ll N^2(\log N)^{-2B},$$

so that

$$\int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\,d\beta = \int_0^1 T(\beta)^3 e(-N\beta)\,d\beta + O(N^2(\log N)^{-2B}).$$

The integral on the right is equal to the number of ways of writing N in the form $N = k_1 + k_2 + k_3$, and this is

$$\tfrac{1}{2}(N-1)(N-2) = \tfrac{1}{2}N^2 + O(N).$$

Hence

$$(5)\quad \int_{-1/Q}^{1/Q} T(\beta)^3 e(-N\beta)\,d\beta = \tfrac{1}{2}N^2 + O(N^2(\log N)^{-2B}).$$

To deal with the sum in (4) we first evaluate Ramanujan's sum $c_q(n)$. Grouping residue classes $a \pmod q$ according to the value of (a, q), we see that

$$\sum_{a=1}^{q} e\left(\frac{an}{q}\right) = \sum_{d|q} \sum_{\substack{a=1 \\ (a,q)=d}}^{q} e\left(\frac{an}{q}\right) = \sum_{d|q} c_{q/d}(n).$$

Here the sum on the left vanishes if q does not divide n, and is equal to q if q divides n; thus by Möbius inversion,

$$c_q(n) = \sum_{\substack{d|n \\ d|q}} d\mu\left(\frac{q}{d}\right). \tag{6}$$

It is now clear that $c_q(n)$ is a multiplicative function of q for any fixed n. Let p^α be the highest power of p dividing n. Then $c_{p^\beta}(n) = \phi(p^\beta)$ for $\beta \le \alpha$, $c_{p^{\alpha+1}}(n) = -p^\alpha$, and $c_{p^\beta}(n) = 0$ for $\beta > \alpha + 1$. Hence

$$c_q(n) = \frac{\mu(q/(n, q))\phi(q)}{\phi(q/(n, q))}. \tag{7}$$

From the trivial estimate $|c_q(n)| \le \phi(q)$ we see that

$$\sum_{q>P} \frac{\mu(q)}{\phi(q)^3} c_q(N) \ll \sum_{q>P} \frac{1}{\phi(q)^2} \ll (\log N)^{-B+1}.$$

The sum in (4), when extended over all q, can be written as an absolutely convergent product,

$$\begin{aligned}\sum_{q=1}^{\infty} \frac{\mu(q)}{\phi(q)^3} c_q(N) &= \prod_p \left(1 - \frac{c_p(N)}{(p-1)^3}\right) \\ &= \prod_{p|N}\left(1 - \frac{1}{(p-1)^2}\right) \prod_{p\nmid N}\left(1 + \frac{1}{(p-1)^3}\right) \\ &= \mathfrak{S}(N),\end{aligned}$$

so that

$$\sum_{q\le P} \frac{\mu(q)}{\phi(q)^3} c_q(N) = \mathfrak{S}(N) + O((\log N)^{-B+1}).$$

We combine this with (5) in (4) to see that

$$\int_{\mathfrak{M}} S(\alpha)^3 e(-N\alpha)\, d\alpha = \tfrac{1}{2}\mathfrak{S}(N)N^2 + O(N^2(\log N)^{-B+1}). \tag{8}$$

To complete the argument we must show that the minor arcs contribute a smaller amount. We note that

$$\begin{aligned}\left|\int_{\mathfrak{m}} S(\alpha)^3 e(-N\alpha)\, d\alpha\right| &\le \left(\max_{\mathfrak{m}} |S(\alpha)|\right)\int_{\mathfrak{m}} |S(\alpha)|^2\, d\alpha \\ &\le \left(\max_{\mathfrak{m}} |S(\alpha)|\right)\int_0^1 |S(\alpha)|^2\, d\alpha.\end{aligned}$$

This last integral is

$$\sum_{k_1 \le N} \Lambda(k_1) \sum_{k_2 \le N} \Lambda(k_2) \int_0^1 e((k_1 - k_2)\alpha)\, d\alpha = \sum_{k \le N} \Lambda(k)^2 \ll N \log N.$$

Dirichlet's theorem on Diophantine approximation asserts that for any real α and any real number $Q \geq 1$, there is a rational number a/q such that $|\alpha - a/q| \leq 1/qQ$, $1 \leq q \leq Q$, and $(a, q) = 1$. If $q \leq P$, then $\alpha \in \mathfrak{M}(q, a)$; hence if $\alpha \in \mathfrak{m}$, then $P < q \leq Q$. That is, for each $\alpha \in \mathfrak{m}$ we have a/q with $|\alpha - a/q| \leq 1/qQ \leq 1/q^2$, $(a, q) = 1$, and $P < q \leq Q$. Hence by our estimates in §25,

$$S(\alpha) \ll N(\log N)^{-(B/2)+4}$$

for $\alpha \in \mathfrak{m}$, and therefore

$$\int_{\mathfrak{m}} S(\alpha)^3 e(-N\alpha)\, d\alpha \ll N^2(\log N)^{-(B/2)+5}.$$

This with (8) gives the desired result, on taking $B = 2A + 10$.

27

THE LARGE SIEVE

The large sieve was first proposed by Linnik[1] in a short but important paper of 1941. In a subsequent series of papers, Rényi developed the method by adopting a probabilistic attitude. His estimates were not optimal, and in 1965 Roth[2] substantially modified Rényi's approach to obtain an essentially optimal result. Bombieri[3] further refined the large sieve, and used it to describe the distribution of primes in arithmetic progressions; this we shall discuss in the following section.

Rényi's approach to the large sieve concerns an extension of Bessel's inequality. We recall that Bessel's inequality asserts that if $\phi_1, \phi_2, \ldots, \phi_R$ are orthonormal members of an inner product space V over the complex numbers, and if $\xi \in V$, then

$$\sum_{r=1}^{R} |(\xi, \phi_r)|^2 \leq \|\xi\|^2.$$

In number theory we frequently encounter vectors which are not quite orthonormal. Thus, with possible applications in mind, we seek an inequality

$$\sum_{r=1}^{R} |(\xi, \phi_r)|^2 \leq A\|\xi\|^2, \tag{1}$$

valid for all ξ, where A depends on $\phi_1, \ldots, \phi_R$; we hope to find that A is near 1 when the ϕ_r are in some sense nearly orthonormal. Boas[4] has characterized the constant A for which (1) holds: The inequality (1) holds for all ξ if and only if

$$\sum_{\substack{1 \leq r \leq R \\ 1 \leq s \leq R}} u_r \bar{u}_s(\phi_r, \phi_s) \leq A \sum_{r=1}^{R} |u_r|^2 \tag{2}$$

[1] *Dokl. Akad. Nauk SSSR*, **30**, 292–294 (1941).
[2] *Mathematika*, **12**, 1–9 (1965).
[3] *Mathematika*, **12**, 201–225 (1965).
[4] *Amer. J. Math.*, **63**, 361–370 (1941).

for all complex numbers u_r. To see this, suppose first that (2) holds. Then

$$0 \le \left\| \xi - \sum_{r=1}^{R} u_r \phi_r \right\|^2 = \|\xi\|^2 - 2\Re \sum_{r=1}^{R} \bar{u}_r(\xi, \phi_r) + \sum_{r,s} u_r \bar{u}_s(\phi_r, \phi_s),$$

and by (2) this is

$$\le \|\xi\|^2 - 2\Re \sum_{r=1}^{R} \bar{u}_r(\xi, \phi_r) + A \sum_{r=1}^{R} |u_r|^2.$$

We now take $u_r = (\xi, \phi_r)/A$, and then the above simplifies to read

$$0 \le \|\xi\|^2 - \frac{1}{A} \sum_{r=1}^{R} |(\xi, \phi_r)|^2,$$

which gives (1). We note that if the ϕ_r are orthonormal then equality holds in (2) with $A = 1$, and then our argument reduces to the usual proof of Bessel's inequality.

To demonstrate the converse, we assume that (1) holds for all ξ, and we take $\xi = \sum_{r=1}^{R} u_r \phi_r$. Then the left-hand side of (2) is

$$\|\xi\|^2 = \sum_{s=1}^{R} \bar{u}_s(\xi, \phi_s).$$

By Cauchy's inequality this is

$$\le \left(\sum_{s=1}^{R} |u_s|^2 \right)^{\frac{1}{2}} \left(\sum_{s=1}^{R} |(\xi, \phi_s)|^2 \right)^{\frac{1}{2}},$$

and by (1) this is

$$\le A^{\frac{1}{2}} \|\xi\| \left(\sum_{s=1}^{R} |u_s|^2 \right)^{\frac{1}{2}}.$$

We divide both sides by $\|\xi\|$ and square, to see that

$$\|\xi\|^2 \le A \sum_{s=1}^{R} |u_s|^2,$$

which is (2).

A great deal is known concerning bounds for bilinear forms such as (2); we content ourselves with a simple argument which is not always efficient but which suffices here. We note that

$$|u_r \bar{u}_s| \le \tfrac{1}{2}|u_r|^2 + \tfrac{1}{2}|u_s|^2;$$

hence the left-hand side of (2) is

$$\le \sum_{r,s} (\tfrac{1}{2}|u_r|^2 + \tfrac{1}{2}|u_s|^2)|(\phi_r, \phi_s)| = \sum_r |u_r|^2 \sum_{s=1}^{R} |(\phi_r, \phi_s)|$$

$$\le \left(\max_r \sum_{s=1}^{R} |(\phi_r, \phi_s)|\right) \sum_{r=1}^{R} |u_r|^2.$$

Thus (2) holds with

$$A = \max_r \sum_{s=1}^{R} |(\phi_r, \phi_s)|, \tag{3}$$

and we have proved

THEOREM 1. *Let* $\phi_1, \phi_2, \ldots, \phi_R$ *and* ξ *be arbitrary vectors in an inner product space* V *over the complex numbers. Then*

$$\sum_{r=1}^{R} |(\xi, \phi_r)|^2 \le A\|\xi\|^2,$$

where A *is given by* (3).

If the ϕ_r are orthonormal, then $A = 1$ in (3), and we see that the above includes Bessel's inequality as a special case. Moreover, if the inner product matrix $[(\phi_r, \phi_s)]$ is near the identity, then A is near 1.

Rényi applied inequalities such as the above directly to arithmetic sequences. One of Roth's innovations was to begin with exponential sums; this yielded vectors which are more nearly orthogonal. Following Davenport and Halberstam[5], we consider the large sieve to be an inequality of the following kind: Let

$$S(\alpha) = \sum_{n=M+1}^{M+N} a_n e(n\alpha) \tag{4}$$

where M and N are integers, $N > 0$, let $\alpha_1, \alpha_2, \ldots, \alpha_R$ be distinct (mod 1), and let $\delta > 0$ be such that $\|\alpha_r - \alpha_s\| \ge \delta$ for $r \ne$ s. Then for arbitrary a_n,

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \le \Delta \sum_{n=M+1}^{M+N} |a_n|^2. \tag{5}$$

Here Δ is to depend only on N and δ; our first concern is to determine how Δ depends on these two parameters. In passing we note that the value of M is irrelevant, since for any K we can put

$$T(\alpha) = \sum_{n=K+1}^{K+N} a_{M-K+n} e(n\alpha) = e((K - M)\alpha)S(\alpha),$$

[5] *Mathematika*, **13**, 91–96 (1966); **14**, 229–232 (1967).

so that T has frequencies in the range $K + 1 \le n \le K + N$, and $|T(\alpha)| = |S(\alpha)|$.

If $R = 1$ then the situation is particularly simple, for by Cauchy's inequality

$$|S(\alpha)|^2 \le N \sum_{M+1}^{M+N} |a_n|^2. \tag{6}$$

This is best possible, since equality occurs when $a_n = e(-n\alpha)$ for all n. Hence $\Delta \ge N$. On the other hand,

$$\int_0^1 \sum_{r=1}^{R} |S(\alpha_r + \beta)|^2 \, d\beta = R \int_0^1 |S(\beta)|^2 \, d\beta = R \sum_{M+1}^{M+N} |a_n|^2,$$

so that there is a β for which

$$\sum_{r=1}^{R} |S(\alpha_r + \beta)|^2 \ge R \sum_{M+1}^{M+N} |a_n|^2.$$

If $\delta R \le 1$, we can choose R points separated by at least $\delta(\text{mod } 1)$; hence R can be as large as $[\delta^{-1}] \ge \delta^{-1} - 1$ and we see that $\Delta \ge \delta^{-1} - 1$. These considerations show that the following theorem is essentially the best possible.

THEOREM 2. *Let* $S(\alpha)$ *be given by* (4). *Then* (5) *holds with* $\Delta = N + 3\delta^{-1}$.

Proof. We first observe that in view of (6) we may restrict our attention to cases in which $R \ge 2$, so that $\delta \le \frac{1}{2}$. Also, by our remark about the role of M we may assume that $M = -[\frac{1}{2}(N + 1)]$; thus it suffices to show that

$$\sum_{r=1}^{R} \left| \sum_{k=-K}^{K} a_k e(k\alpha_r) \right|^2 \le (2K + 3\delta^{-1}) \sum_{-K}^{K} |a_k|^2$$

for $\delta \le \frac{1}{2}$. We appeal to Theorem 1 with the usual inner product, $(\boldsymbol{\phi}, \boldsymbol{\psi}) = \sum \phi_k \bar{\psi}_k$, taking $\boldsymbol{\xi} = \{a_k b_k^{-\frac{1}{2}}\}_{k=-K}^{K}$ and

$$\boldsymbol{\phi}_r = \{b_k^{\frac{1}{2}} e(-k\alpha_r)\}_{-\infty}^{+\infty}.$$

Here the b_k are nonnegative, and strictly positive for $-K \le k \le K$. Then by Theorem 1,

$$\sum_{r=1}^{R} \left| \sum_{-K}^{K} a_k e(k\alpha_r) \right|^2 \le A \sum_{-K}^{K} |a_k|^2 b_k^{-1},$$

where

$$A = \max_r \sum_{s=1}^{R} |B(\alpha_r - \alpha_s)|;$$

here

$$B(\alpha) = \sum_{-\infty}^{+\infty} b_k e(k\alpha).$$

It now suffices to choose nonnegative b_k such that $b_k \geq 1$ for $-K \leq k \leq K$, and such that

$$\sum_{s=1}^{R} |B(\alpha_r - \alpha_s)| \leq 2K + 3\delta^{-1} \tag{7}$$

for all r. If we were to take $b_k = 1$ for $-K \leq k \leq K$, $b_k = 0$ otherwise, we would obtain the inferior estimate

$$\sum_{s=1}^{R} |B(\alpha_r - \alpha_s)| \leq 2K + O(\delta^{-1} \log \delta^{-1}).$$

To obtain a sharper estimate we take smoother b_k, namely

$$b_k = \begin{cases} 1 & \text{if } |k| \leq K, \\ 1 - (|k| - K)/L & \text{if } K \leq |k| \leq K + L, \\ 0 & \text{if } |k| \geq K + L, \end{cases}$$

where L is a positive integer to be selected later. To write $B(\alpha)$ in closed form we appeal to the identity

$$\sum_{|j| \leq J} (J - |j|)e(j\alpha) = \left| \sum_{j=1}^{J} e(j\alpha) \right|^2 = \left(\frac{\sin \pi J\alpha}{\sin \pi\alpha} \right)^2,$$

firstly with $J = K + L$ and secondly with $J = K$, for by subtraction we then find that

$$B(\alpha) = \frac{1}{L}((\sin \pi(K + L)\alpha)^2 - (\sin \pi K\alpha)^2)(\sin \pi\alpha)^{-2}.$$

Hence $B(0) = 2K + L$, and

$$|B(\alpha)| \leq \frac{1}{L}(\sin \pi\alpha)^{-2} \leq (4L\|\alpha\|^2)^{-1},$$

so that

$$\sum_{s=1}^{R} |B(\alpha_r - \alpha_s)| \leq 2K + L + 2\sum_{h=1}^{\infty} (4Lh^2\delta^2)^{-1}.$$

To evaluate this last term it is useful to know that

$$\sum_{h=1}^{\infty} h^{-2} = \zeta(2) = \frac{\pi^2}{6}.$$

However, for our purposes it is sufficient to note that

$$\sum_{h=1}^{\infty} h^{-2} < 1 + \int_1^{\infty} u^{-2}\,du = 2,$$

by the integral test. Hence

$$\sum_{s=1}^{R} |B(\alpha_r - \alpha_s)| \le 2K + L + \frac{1}{L\delta^2}\cdot$$

We now let L be the least integer $\ge \delta^{-1}$, for then the above is

$$\ll 2K + \delta^{-1} + 1 + \delta^{-1} \le 2K + 3\delta^{-1},$$

since $\delta \le \frac{1}{2}$. Thus we have (7), and the proof is complete.

A. Selberg chose the b_k more carefully, and obtained the sharper value $\Delta = N + \delta^{-1} - 1$; this and other refinements are found in the survey article of Montgomery[6].

Gallagher[7] has devised a different approach to the large sieve; his method gives $\Delta = \pi N + \delta^{-1}$ which is sharper than Theorem 2 when $N\delta$ is small. We do not need his results, but we describe his method as it is very flexible, and can be used to advantage in other contexts. If f has a continuous first derivative in $[0, 1]$ then

$$f(x) = \int_0^1 f(u)\,du + \int_0^x uf'(u)\,du + \int_x^1 (u-1)f'(u)\,du$$

for $0 \le x \le 1$, as we may verify by integrating by parts. Hence

$$f(\tfrac{1}{2}) \le \int_0^1 |f(u)| + \tfrac{1}{2}|f'(u)|\,du,$$

and in general

$$|f(x)| \le \int_0^1 |f(u)| + |f'(u)|\,du$$

for $0 \le x \le 1$. After a change of variables the first inequality takes the form

$$|f(\alpha)| \le \int_{\alpha-\frac{1}{2}\delta}^{\alpha+\frac{1}{2}\delta} \frac{1}{\delta}|f(u)| + \tfrac{1}{2}|f'(u)|\,du.$$

We take $f(\alpha) = S(\alpha)^2$ and sum, to see that

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \le \sum_{r=1}^{R} \int_{\alpha_r-\frac{1}{2}\delta}^{\alpha_r+\frac{1}{2}\delta} \frac{1}{\delta}|S(\alpha)|^2 + |S(\alpha)S'(\alpha)|\,d\alpha.$$

[6] *Bull. Amer. Math. Soc.*, **84**, 547–567 (1978).
[7] *Mathematika*, **14**, 14–20 (1967).

The intervals of integration are nonoverlapping and the integrand is nonnegative, so the above is

$$\leq \frac{1}{\delta}\int_0^1 |S(\alpha)|^2\,d\alpha + \int_0^1 |S(\alpha)S'(\alpha)|\,d\alpha.$$

By Parseval's identity the first integral is $\sum_{M+1}^{M+N}|a_n|^2$; this is easily verified by expanding and integrating term-by-term. By Cauchy's inequality, the second integral is

$$\leq \left(\int_0^1 |S(\alpha)|^2\,d\alpha\right)^{\frac{1}{2}}\left(\int_0^1 |S'(\alpha)|^2\,d\alpha\right)^{\frac{1}{2}}.$$

Again by Parseval's identity, this is

$$\left(\sum_{M+1}^{M+N}|a_n|^2\right)^{\frac{1}{2}}\left(\sum_{M+1}^{M+N}|2\pi i n a_n|^2\right)^{\frac{1}{2}}.$$

Without loss of generality we may suppose that $M = -[\frac{1}{2}(N+1)]$, so that $|n| \leq \frac{1}{2}N$ for $M+1 \leq n \leq M+N$. Then the above is

$$\leq \pi N \sum_{M+1}^{M+N}|a_n|^2,$$

and we have the large sieve with $\Delta = \delta^{-1} + \pi N$.

In our applications of the large sieve we shall take the points α_r to be the Farey fractions a/q, $(a, q) = 1$, $q \leq Q$. If a/q and a'/q' are two distinct such fractions, then

$$\left\|\frac{a}{q} - \frac{a'}{q'}\right\| = \left\|\frac{aq' - a'q}{qq'}\right\| \geq \frac{1}{qq'} \geq \frac{1}{Q^2};$$

hence we can apply the large sieve with $\delta = Q^{-2}$ to obtain the inequality

$$\sum_{q\leq Q}\sum_{\substack{a=1\\(a,q)=1}}^{q}\left|S\left(\frac{a}{q}\right)\right|^2 \leq (N + 3Q^2)\sum_{M+1}^{M+N}|a_n|^2. \tag{8}$$

We now use the above result to formulate the large sieve in the manner of Rényi. Let $\mathcal{N}$ be a set of Z integers in the interval $M+1 \leq n \leq M+N$, and let $Z(q, h)$ denote the number of these integers which are congruent to $h \pmod{q}$. Clearly

$$\sum_{h=1}^{q} Z(q, h) = Z,$$

so that the average of $Z(q, h)$ is Z/q. Rényi considered the mean square error, i.e., the "variance"

$$V(q) = \sum_{h=1}^{q} \left(Z(q, h) - \frac{Z}{q}\right)^2.$$

From the large sieve we find that the numbers $V(p)$ are on average small; we find that

$$\sum_{p \le Q} pV(p) \le (N + 3Q^2)Z. \tag{9}$$

To see this we let a_n be the characteristic function of the set $\mathscr{N}$, so that

$$S(\alpha) = \sum_{n \in \mathscr{N}} e(n\alpha).$$

Then

$$\sum_{a=1}^{q} \left| S\left(\frac{a}{q}\right) \right|^2 = \sum_{m \in \mathscr{N}} \sum_{n \in \mathscr{N}} \sum_{a=1}^{q} e\left(\frac{a(m-n)}{q}\right).$$

The innermost sum is $= q$ or 0, according as $m \equiv n \pmod q$ or not; hence the above is

$$q \sum_{\substack{m \in \mathscr{N} \\ m \equiv n \pmod q}} \sum_{n \in \mathscr{N}} 1 = q \sum_{h=1}^{q} Z(q, h)^2.$$

Thus when we expand the square in $V(q)$, we see that

$$\begin{aligned} qV(q) &= q \sum_{h=1}^{q} Z(q, h)^2 - 2Z \sum_{h=1}^{q} Z(q, h) + Z^2 \\ &= \sum_{a=1}^{q} \left| S\left(\frac{a}{q}\right) \right|^2 - Z^2. \end{aligned}$$

But $S(0) = Z$, so that

$$qV(q) = \sum_{a=1}^{q-1} \left| S\left(\frac{a}{q}\right) \right|^2,$$

and consequently by (8),

$$\sum_{p \le Q} pV(p) = \sum_{p \le Q} \sum_{a=1}^{p-1} \left| S\left(\frac{a}{p}\right) \right|^2 \le (N + 3Q^2)Z.$$

Using (9) we can now present the large sieve as a sieve in the elementary sense. Suppose that from the interval $M + 1 \le n \le M + N$ we remove several arithmetic progressions, and we let $\mathscr{N}$ denote the remaining set. For example, suppose that we have removed those numbers congruent to $h \pmod q$. Then $Z(q, h)$ is not near Z/q as

would normally be the case, but instead $Z(q, h) = 0$. If this occurs for many $h \pmod q$, then $V(q)$ is large, and if $V(p)$ is large for many primes p, then Z is small. More specifically, we have

THEOREM 3. *Let $\mathcal{N}$ be a set of Z integers in the interval $M + 1 \le n \le M + N$. Let $\mathcal{P}$ be a set of P prime numbers p, with $p \le Q$ for all $p \in \mathcal{P}$. Let $0 < \tau < 1$, and suppose that $Z(p, h) = 0$ for at least τp values of h (mod p), for all $p \in \mathcal{P}$. Then*

$$Z \le \frac{N + 3Q^2}{\tau P}.$$

To see this we note that if $p \in \mathcal{P}$, then $V(p) \ge \tau p(Z/p)^2$, so that by (9),

$$\tau P Z^2 \le (N + 3Q^2)Z.$$

This gives the desired bound.

To appreciate the strength of this bound, suppose that $\mathcal{N}$ is the set of squares in the interval $1 \le n \le N$, let $Q = N^{\frac{1}{2}}$, and let $\mathcal{P}$ be the set of odd primes $p \le N^{\frac{1}{2}}$. Then $Z(p, h) = 0$ for quadratic non-residues $h \pmod p$, so that $Z(p, h) = 0$ for at least $\frac{1}{2}(p - 1)$ values of h. Hence $\tau = \frac{1}{3}$ and $P \sim 2N^{\frac{1}{2}}/\log N$, and we obtain the bound $Z \ll N^{\frac{1}{2}} \log N$, which is not far from the truth, $Z \sim N^{\frac{1}{2}}$.

To derive (9) from (8) we used only prime moduli. By taking more care we can use composite moduli as well, and thus obtain a sharper bound. This was first done by Bombieri and Davenport[8] in a special case, and by Montgomery[9] in general; the result is that if $\mathcal{N}$ is a set of Z members in the interval $M + 1 \le n \le M + N$, and if $\omega(p)$ is the number of $h \pmod p$ for which $Z(p, h) = 0$, then

$$Z \le \frac{N + 3Q^2}{L},$$

where

$$L = \sum_{q \le Q} \mu^2(q) \prod_{p \mid q} \frac{\omega(p)}{p - \omega(p)}.$$

The large sieve, in the form of inequality (8), is useful also in estimating averages of character sums, as was first observed by Rényi[10]. Gallagher[11] found the following elegant formulation.

[8] *Abh. aus Zahlentheorie und Analysis zur Erinnerung an Edmund Landau*, VEB Deutsch. Verlag Wiss., Berlin, 1968, 11–22.

[9] *J. London Math. Soc.*, **43**, 93–98 (1968).

[10] *Izv. Akad. Nauk SSSR Ser. Mat.*, **12**, 57–78 (1948); *Amer. Math. Soc. Transl.*, (2) **19**, 299–321 (1962).

[11] *Mathematika*, **14**, 14–20 (1967).

THEOREM 4. *Let* χ *be a character* (mod q), *and put* $T(\chi) = \sum_{M+1}^{M+N} a_n\chi(n)$. *Then for any* $Q \geq 1$,

$$\sum_{q\leq Q} \frac{q}{\phi(q)} \sum_{\chi}{}^{*} |T(\chi)|^2 \leq (N + 3Q^2) \sum_{M+1}^{M+N} |a_n|^2.$$

Here $\sum_\chi^*$ *denotes a sum over all primitive characters* χ (mod q).

It suffices to show that

$$\sum_{\chi}{}^{*} |T(\chi)|^2 \leq \frac{\phi(q)}{q} \sum_{\substack{a=1\\(a,q)=1}}^{q} \left|S\left(\frac{a}{q}\right)\right|^2, \tag{10}$$

for then the result follows from (8). To establish (10), we recall fron §9 that if χ is primitive (mod q), then

$$\chi(n) = \frac{1}{\tau(\bar{\chi})} \sum_{a=1}^{q} \bar{\chi}(a) e\left(\frac{an}{q}\right)$$

for all n. On multiplying both sides by a_n and summing, we see that

$$T(\chi) = \frac{1}{\tau(\bar{\chi})} \sum_{a=1}^{q} \bar{\chi}(a) S\left(\frac{a}{q}\right).$$

As $|\tau(\bar{\chi})| = q^{\frac{1}{2}}$ for primitive χ, we find that

$$\sum_{\chi}{}^{*} |T(\chi)|^2 = \frac{1}{q} \sum_{\chi}{}^{*} \left| \sum_{a=1}^{q} \bar{\chi}(a) S\left(\frac{a}{q}\right)\right|^2.$$

The right-hand side is increased if we drop the condition that χ be primitive, and

$$\begin{aligned}\sum_{\chi} \left| \sum_{a=1}^{q} \bar{\chi}(a) S\left(\frac{a}{q}\right)\right|^2 &= \sum_{a=1}^{q} \sum_{b=1}^{q} S\left(\frac{a}{q}\right) \overline{S\left(\frac{b}{q}\right)} \sum_{\chi} \bar{\chi}(a)\chi(b) \\ &= \phi(q) \sum_{\substack{a=1\\(a,q)=1}}^{q} \left|S\left(\frac{a}{q}\right)\right|^2.\end{aligned}$$

Thus we have (10), and the proof is complete.

28

BOMBIERI'S THEOREM

Rényi used the large sieve to show that prime numbers are well distributed in arithmetic progressions (mod q) for most q; his rather complicated result allowed him to show that every large even number is representable in the form

$$p + p_1 p_2 \cdots p_r,$$

where r is bounded by some absolute constant. The subsequent refinements of Bombieri[1] and A. I. Vinogradov[2] enable one to take $r = 3$, and recently Chen[3] has added an ingenious new idea to obtain $r = 2$.

We now develop Bombieri's elegant estimate, without pursuing its applications. For brevity we put

$$E(x; q, a) = \psi(x; q, a) - \frac{x}{\phi(q)}$$

for $(a, q) = 1$, we let

$$E(x; q) = \max_{\substack{a \\ (a,q)=1}} |E(x; q, a)|,$$

and

$$E^*(x, q) = \max_{y \le x} E(y, q).$$

We prove that $E^*(x, q)$ is significantly smaller than $x/\phi(q)$ for most $q \le x^{\frac{1}{2}}(\log x)^{-A}$.

THEOREM. *Let $A > 0$ be fixed. Then*

$$\sum_{q \le Q} E^*(x, q) \ll x^{\frac{1}{2}} Q (\log x)^5 \tag{1}$$

provided that $x^{\frac{1}{2}}(\log x)^{-A} \le Q \le x^{\frac{1}{2}}$.

[1] *Mathematika*, **12**, 201–225 (1965).

[2] *Izv. Akad. Nauk SSSR Ser. Mat.*, **29**, 903–934 (1965); **30**, 719–720 (1966).

[3] *Sci. Sinica*, **16**, 157–176 (1973); see also Chapter 11 of Halberstam and Richert, *Sieve Methods*, Academic Press, London (1974).

To assess the strength of this bound we note that there are at most $y/q + 1$ integers $n \le y$, $n \equiv a(\text{mod } q)$, and hence $\psi(y; q, a) \ll xq^{-1} \log x$ for $q \le x, y \le x$, so that $E^*(x, q) \ll xq^{-1} \log x$ for $q \le x$. Consequently the bound

$$\sum_{q \le Q} E^*(x, q) \ll x(\log x)^2$$

is trivial for $Q \le x$. On the other hand, from (1) we see that if $Q = x^{\frac{1}{2}}(\log x)^{-2B-6}$, then

$$\psi(x; q, a) = \frac{x}{\phi(q)}(1 + O((\log x)^{-B}))$$

for all reduced residue classes a (mod q), and for all $q \le Q$ with the possible exception of at most $Q(\log x)^{-B}$ values of q.

Halberstam has conjectured that

$$\sum_{q \le x^{1-\varepsilon}} E^*(x, q) \ll x(\log x)^{-A}$$

for any fixed positive A and ε; such a strengthening of Bombieri's theorem would have important consequences.

Our proof of the theorem falls in two parts. First we use our estimates of §22 to show that the theorem follows from the bound

$$(2) \quad \sum_{q \le Q} \frac{q}{\phi(q)} \sum_{\chi}{}^* \max_{y \le x} |\psi(y, \chi)| \ll (x + x^{\frac{5}{6}}Q + x^{\frac{1}{2}}Q^2)(\log Qx)^4,$$

which is valid for all $x \ge 1$, $Q \ge 1$. Then we establish (2) by combining the large sieve with the method of §24.

We recall that

$$\psi(y; q, a) = \frac{1}{\phi(q)} \sum_{\chi} \bar{\chi}(a)\psi(y, \chi).$$

From $\psi(y, \chi_0)$ we wish to subtract the main term y; accordingly we put

$$\psi'(y, \chi) = \begin{cases} \psi(y, \chi) & \text{if } \chi \ne \chi_0, \\ \psi(y, \chi_0) - y & \text{if } \chi = \chi_0. \end{cases}$$

Then

$$\psi(y; q, a) - \frac{y}{\phi(q)} = \frac{1}{\phi(q)} \sum_{\chi} \bar{\chi}(a)\psi'(y, \chi),$$

and hence

$$|E(y; q, a)| \le \frac{1}{\phi(q)} \sum_{\chi} |\psi'(y, \chi)|.$$

As this estimate is independent of a, we see that $E(y; q)$ satisfies the same bound. If $\chi \pmod q$ is induced by $\chi_1 \pmod{q_1}$ then $\psi'(y, \chi)$ and $\psi'(y, \chi_1)$ are nearly equal, for

$$\begin{aligned}\psi'(y, \chi_1) - \psi'(y, \chi) &= \sum_{\substack{p^k \le y \\ p \mid q}} \chi_1(p^k)\log p \\ &\ll \sum_{p \mid q} \left[\frac{\log y}{\log p}\right]\log p \\ &\ll (\log y)\sum_{p \mid q} \log p \ll (\log qy)^2.\end{aligned}$$

Hence

$$E(y, q) \ll (\log qy)^2 + \frac{1}{\phi(q)}\sum_{\chi} |\psi'(y, \chi_1)|,$$

and thus

$$E^*(x, q) \ll (\log qx)^2 + \frac{1}{\phi(q)}\sum_{\chi} \max_{y \le x} |\psi'(y, \chi_1)|.$$

We now combine all contributions made by an individual primitive character. A primitive character $\chi \pmod q$ induces characters to moduli which are multiples of q; hence the left-hand side of (1) is

$$\ll Q(\log Qx)^2 + \sum_{q \le Q} \sideset{}{^*}\sum_{\chi} \max_{y \le x} |\psi'(y, \chi)| \left(\sum_{k \le Q/q} \frac{1}{\phi(kq)}\right).$$

Here the first term is negligible. As for the second term, we note that $\phi(kq) \ge \phi(k)\phi(q)$, so that

$$\sum_{k \le z} \frac{1}{\phi(kq)} \le \frac{1}{\phi(q)} \sum_{k \le z} \frac{1}{\phi(k)}.$$

Moreover,

$$\begin{aligned}\sum_{k \le z} \frac{1}{\phi(k)} &\le \prod_{p \le z}\left(1 + \frac{1}{p-1} + \frac{1}{p(p-1)} + \cdots\right) \\ &= \prod_{p \le z}\left(1 - \frac{1}{p}\right)^{-1}\left(1 + \frac{1}{p(p-1)}\right) \\ &\ll \log z.\end{aligned}$$

Hence the second term above is

$$\ll (\log x) \sum_{q \le Q} \frac{1}{\phi(q)} \sideset{}{^*}\sum_{\chi} \max_{y \le x} |\psi'(y, \chi)|,$$

and so it suffices to show that

$$\text{(3)} \qquad \sum_{q\le Q}\frac{1}{\phi(q)}\sum_{\chi}{}^{*}\max_{y\le x}|\psi'(y,\chi)| \ll x^{\frac{1}{2}}Q(\log x)^4$$

for $x^{\frac{1}{2}}(\log x)^{-A} \le Q \le x^{\frac{1}{2}}$. We now consider large and small values of q separately. From (2) we see that

$$\sum_{U<q\le 2U}\frac{1}{\phi(q)}\sum_{\chi}{}^{*}\max_{y\le x}|\psi(y,\chi)| \ll \left(\frac{x}{U} + x^{\frac{5}{6}} + x^{\frac{1}{2}}U\right)(\log Ux)^4.$$

By summing this over $U = 2^k$ for an appropriate range of k, we see that

$$\sum_{Q_1<q\le Q}\frac{1}{\phi(q)}\sum_{\chi}{}^{*}\max_{y\le x}|\psi(y,\chi)| \le \left(\frac{x}{Q_1} + x^{\frac{5}{6}}\log Q + x^{\frac{1}{2}}Q\right)(\log Qx)^4.$$

This is acceptable in (3) if $Q_1 = (\log x)^A$. If χ is a primitive character (mod q), $q \le (\log x)^A$, $y \le x$, then by estimate (3) of §22,

$$\psi'(y,\chi) \ll x(\log x)^{-2A},$$

and hence the contribution of $q \le (\log x)^A$ in (3) is $\ll x(\log x)^{-A}$, which is also acceptable. Thus the theorem follows from (2).

We now prove the estimate (2). In §24 we observed that our method of estimating $\sum_{n\le N} f(n)\Lambda(n)$ fails if f is multiplicative; in particular we are not able to bound $\psi(x,\chi)$ by this method. Nevertheless we can use the method to bound an average of $|\psi(x,\chi)|$ over various χ, by using the large sieve. More precisely, we use the large sieve in the form of the inequality

$$\text{(4)} \qquad \sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi}{}^{*}\left|\sum_{M+1}^{M+N} a_n\chi(n)\right|^2 \ll (N+Q^2)\sum_{M+1}^{M+N}|a_n|^2,$$

to show that

$$\text{(5)} \qquad \sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi}{}^{*}\max_{u}\left|\sum_{\substack{1\le m\le M\\ mn\le u}}\sum_{1\le n\le N} a_m b_n\chi(mn)\right|$$

$$\ll (M+Q^2)^{\frac{1}{2}}(N+Q^2)^{\frac{1}{2}}\left(\sum_{1\le m\le M}|a_m|^2\right)^{\frac{1}{2}}\left(\sum_{1\le n\le N}|b_n|^2\right)^{\frac{1}{2}}\log 2MN;$$

we use this in the method of §24.

To derive (5) we first note that by (4) and Cauchy's inequality,

$$\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi}{}^{*}\left|\sum_{m=1}^{M}\sum_{n=1}^{N}a_m b_n\chi(mn)\right|$$

$$\le\left(\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi}{}^{*}\left|\sum_{m=1}^{M}a_m\chi(m)\right|^2\right)^{\frac12}\left(\sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi}{}^{*}\left|\sum_{n=1}^{N}b_n\chi(n)\right|^2\right)^{\frac12}$$

$$\ll (M+Q^2)^{\frac12}(N+Q^2)^{\frac12}\left(\sum_{m=1}^{M}|a_m|^2\right)^{\frac12}\left(\sum_{n=1}^{N}|b_n|^2\right)^{\frac12}. \tag{6}$$

To introduce the condition $mn \le u$ we appeal to the Lemma of §17 (with c tending to 0), from which we see that if $T > 0$, $\beta > 0$, and α is real, then

$$\int_{-T}^{T} e^{it\alpha}\frac{\sin t\beta}{t}\,dt = \begin{cases}\pi + O(T^{-1}(\beta-|\alpha|)^{-1}) & \text{if } |\alpha|\le\beta,\\ O(T^{-1}(|\alpha|-\beta)^{-1}) & \text{if } |\alpha|>\beta.\end{cases}$$

Putting $\beta = \log u$, we find that

$$\sum_{\substack{m=1\\ mn\le u}}^{M}\sum_{n=1}^{N}a_m b_n\chi(mn) = \int_{-T}^{T}A(t,\chi)B(t,\chi)\frac{\sin(t\log u)}{\pi t}\,dt$$

$$+\,O\left(T^{-1}\sum_{m,n}|a_m b_n|\left|\log\frac{mn}{u}\right|^{-1}\right),$$

where

$$A(t,\chi) = \sum_{m=1}^{M}a_m\chi(m)m^{-it}, \qquad B(t,\chi) = \sum_{n=1}^{N}b_n\chi(n)n^{-it}.$$

Without loss of generality we may assume that u is of the form $u = k + \frac12$, where k is an integer, $0 \le k \le MN$. Then

$$\left|\log\frac{mn}{u}\right| \gg \frac{1}{u} \gg \frac{1}{MN},$$

and

$$\sin(t\log u) \ll \min(1, |t|\log 2MN),$$

so that the right-hand side above is

$$\ll \int_{-T}^{T}|A(t,\chi)B(t,\chi)|\min\left(\frac{1}{|t|}, \log 2MN\right)dt + \frac{MN}{T}\sum_{m,n}|a_m b_n|.$$

We now apply (6) to the first term, and Cauchy's inequality to the second, in order to see that the left-hand side of (5) is

$$\ll (M + Q^2)^{\frac{1}{2}}(N + Q^2)^{\frac{1}{2}}\left(\sum_{m=1}^{M} |a_m|^2\right)^{\frac{1}{2}}$$

$$\times \left(\sum_{n=1}^{N} |b_n|^2\right)^{\frac{1}{2}} \int_{-T}^{T} \min\left(\left|\frac{1}{t}\right|, \log 2MN\right) dt$$

$$+ M^{\frac{1}{2}}N^{\frac{1}{2}}Q^2T^{-1}\left(\sum_{m=1}^{M} |a_m|^2\right)^{\frac{1}{2}}\left(\sum_{n=1}^{N} |b_n|^2\right)^{\frac{1}{2}}.$$

With $T = (MN)^{\frac{3}{2}}$, (5) now follows.

If $Q^2 > x$, then (2) follows from (5) on taking $M = 1$, $a_1 = 1$, $b_n = \Lambda(n)$, $N = x$. We now assume that $Q^2 \leq x$, and prove (2) using the identity of §24. We have

$$\psi(y, \chi) = S_1 + S_2 + S_3 + S_4,$$

where

$$S_1 = \sum_{n \leq U} \Lambda(n)\chi(n) \ll U, \tag{7}$$

$$S_2 = -\sum_{t \leq UV} \left(\sum_{\substack{t = md \\ m \leq U \\ d \leq V}} \mu(d)\Lambda(m)\right) \sum_{r \leq y/t} \chi(rt), \tag{8}$$

$$S_3 \ll (\log y) \sum_{d \leq V} \max_{w} \left| \sum_{w \leq h \leq y/d} \chi(h) \right|, \tag{9}$$

and

$$S_4 = \sum_{U < m \leq y/V} \Lambda(m) \sum_{V < k \leq y/m} \left(\sum_{\substack{d | k \\ d \leq V}} \mu(d)\right)\chi(mk). \tag{10}$$

Here y depends on χ, $y \leq x$, but we shall choose U and V later as functions of Q and x only.

To treat S_4 we first note that by (5),

$$\sum_{q \leq Q} \frac{q}{\phi(q)} \sum_{\chi}{}^{*} \max_{y \leq x} \left| \sum_{\substack{U < m \leq y/V \\ M < m \leq 2M}} \Lambda(m) \sum_{V < k \leq y/m} \left(\sum_{\substack{d | k \\ d \leq V}} \mu(d)\right)\chi(mk) \right|$$

$$\ll (Q^2 + M)^{\frac{1}{2}}\left(Q^2 + \frac{x}{M}\right)^{\frac{1}{2}}\left(\sum_{M}^{2M} \Lambda(m)^2\right)^{\frac{1}{2}}\left(\sum_{k \leq x/M} d(k)^2\right)^{\frac{1}{2}} \log x$$

$$\ll (Q^2x^{\frac{1}{2}} + QxM^{-\frac{1}{2}} + Qx^{\frac{1}{2}}M^{\frac{1}{2}} + x)(\log x)^3. \tag{11}$$

Here we have used the elementary estimates for $\sum \Lambda(m)^2$ and $\sum d(k)^2$ which we proved in §24. We sum (11) over $M = 2^k$ for $\frac{1}{2}U < 2^k \le x/V$, and thus find that

$$(12) \qquad \sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi}{}^{*} \max_{y\le x}|S_4| \ll (Q^2x^{\frac12} + QxU^{-\frac12} + QxV^{-\frac12} + x)(\log x)^4.$$

To treat S_2 we consider two ranges of t, by writing

$$S_2 = \sum_{t\le UV} = \sum_{t\le U} + \sum_{U<t\le UV} = S_2' + S_2''.$$

We deal with S_2'' exactly as we did with S_4, and we find that

$$(13) \qquad \sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi}{}^{*} \max_{y\le x}|S_2''| \ll (Q^2x^{\frac12} + QxU^{-\frac12} + Qx^{\frac12}U^{\frac12}V^{\frac12} + x)(\log x)^2.$$

On the other hand,

$$S_2' \ll (\log U)\sum_{t\le U}\left|\sum_{r\le y/t}\chi(r)\right|,$$

and by the Pólya–Vinogradov inequality of §23 we see that

$$S_2' \ll q^{\frac12}U(\log qU)^2$$

uniformly for $y \le x$. However, this applies only when $q > 1$; for $q = 1$ we have the trivial bound

$$S_2' \ll x(\log xU)^2.$$

On combining these estimates we find that

$$(14) \qquad \sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi}{}^{*} \max_{y\le x}|S_2'| \ll (Q^{\frac52}U + x)(\log Ux)^2.$$

We treat S_3 as we did S_2', and find that

$$(15) \qquad \sum_{q\le Q} \frac{q}{\phi(q)} \sum_{\chi}{}^{*} \max_{y\le x}|S_3| \ll (Q^{\frac52}V + x)(\log Vx)^2.$$

On combining estimates (7), (12)–(15), we see that the left-hand side of (2) is

$$\ll (Q^2x^{\frac12} + x + QxU^{-\frac12} + QxV^{-\frac12} + U^{\frac12}V^{\frac12}Qx^{\frac12} + Q^{\frac52}U + Q^{\frac52}V)(\log xUV)^4.$$

If we allow U and V to vary in such a way that the product UV is fixed, we see that the above is minimized by taking $U = V$. If

$x^{\frac{1}{3}} \le Q \le x^{\frac{1}{2}}$, then the terms involving U are minimized by taking $U = x^{\frac{2}{3}}Q^{-1}$, and then their contribution is

$$\ll Q^{\frac{3}{2}}x^{\frac{2}{3}} \ll Q^2x^{\frac{1}{2}}.$$

If $1 \le Q \le x^{\frac{1}{3}}$, then the terms involving U are minimized by taking $U = x^{\frac{1}{3}}$, and then their contribution is $\ll x^{\frac{5}{6}}Q$. Hence we have (2), and the proof is complete.

We have followed here Vaughan's proof[4] of Bombieri's theorem, which differs significantly from the approach used previously by Rényi and Bombieri. They used the large sieve to estimate the number of zeros of L functions in various rectangles, and then they derived an estimate corresponding to (2) by means of the explicit formulae of §19. Let $N(\sigma, T, \chi)$ denote the number of zeros ρ of $L(s, \chi)$ in the rectangle $\sigma \le \beta \le 1, |\gamma| \le T$. Bombieri[5] proved that

$$\sum_{q \le Q} \sum_{\chi}{}^* N(\sigma, T, \chi) \ll T(Q^2 + QT)^{4(1-\sigma)/(3-2\sigma)}(\log QT)^{10};$$

this was subsequently improved by Montgomery[6] (see also Bombieri[7]). Gallagher[8] proved Bombieri's theorem without discussing zeros, by applying the Mellin transform to the identity

$$-\frac{L'}{L} = -\frac{L'}{L}(1 - LG)^2 - 2L'G + L'LG^2.$$

Vaughan[9] found that it was more efficient to use the identity

$$(16) \qquad -\frac{L'}{L} = F - LFG - L'G + \left(-\frac{L'}{L} - F\right)(1 - LG);$$

he showed that

$$\sum_{q \le Q} \frac{q}{\phi(q)} \sum_{\chi}{}^* \max_{y \le x} |\psi(y, \chi)| \ll (Q^2x^{\frac{1}{2}} + Q^{\frac{5}{4}}x^{\frac{3}{4}} + x)(\log Qx)^4.$$

Then Vaughan discovered that the identity (16) could be used to provide a new form of Vinogradov's method; this permitted us to derive the sharp estimate (2) by essentially elementary means.

[4] To appear in the Turán memorial volume of *Acta Arithmetica*, **37**, 111–115 (1980).

[5] *Mathematika*, **12**, 201–225 (1965).

[6] *Topics in Multiplicative Number Theory*, Springer-Verlag, Berlin (1971) Chapter 12.

[7] *Le grand crible dans la théorie analytique des nombres*, Astérisque No. 18, Soc. Math. France, Paris, 1974.

[8] *Mathematika*, **15**, 1–6 (1968).

[9] *J. London Math. Soc.*, (2) **10**, 153–162 (1975).

29

AN AVERAGE RESULT

We now consider the mean square error in the prime number theorem for arithmetic progressions. Work in this direction was initiated by Barban[1], and by Davenport and Halberstam[2]. Their results were sharpened by Gallagher[3], who showed that

$$(1)\qquad \sum_{q\le Q}\ \sum_{\substack{a=1\\(a,q)=1}}^{q}\left(\psi(x;q,a)-\frac{x}{\phi(q)}\right)^2 \ll xQ\log x$$

for $x(\log x)^{-A}\le Q\le x$; here $A>0$ is fixed. This estimate is best possible, for Montgomery[4] has shown that the left-hand side is $\sim Qx\log x$ for Q in the stated range. Moreover, Hooley[5] has shown that (1) can be combined with some of Montgomery's ideas to give, in a simple way, a very precise asymptotic estimate.

The estimate (1) differs from Bombieri's theorem of §28 in that we have a much longer range of q, and we consider a mean over residue classes instead of the maximum. We again use the large sieve, but now the proof is simpler than in the case of Bombieri's theorem. In fact, by the large sieve in the form of Theorem 4 of §27, with $a_n = \Lambda(n)$, we have

$$(2)\qquad \sum_{q\le Q}\frac{q}{\phi(q)}\sum_{\chi}{}^{*}|\psi(x,\chi)|^2 \ll (x+Q^2)x\log x,$$

since $\Sigma_{n\le x}\Lambda(n)^2\ll x\log x$. We now derive (1) from (2) in much the same way that we derived (1) from (2) in the previous section. As in that argument,

$$(3)\qquad \psi(x;q,a)-\frac{x}{\phi(q)}=\frac{1}{\phi(q)}\sum_{\chi}\bar{\chi}(a)\psi'(x,\chi).$$

[1] *Dokl. Akad. Nauk UzSSR*, **1964**, No. 5, 5–7.
[2] *Michigan Math. J.*, **13**, 485–489 (1966); **15**, 505 (1968).
[3] *Mathematika*, **14**, 14–20 (1967).
[4] *Michigan Math. J.*, **17**, 33–39 (1970).
[5] *J. Reine Angew. Math.*, **274/275**, 206–223 (1975).

We now form the square of the modulus of both sides, and sum over a. We expand the right-hand side and take the sum over a inside, to see that

$$\sum_{\substack{a=1\\(a,q)=1}}^{q} \Big|\sum_{\chi} \bar{\chi}(a)\psi'(x,\chi)\Big|^2 = \phi(q)\sum_{\chi}|\psi'(x,\chi)|^2$$

since

$$\sum_{\substack{a=1\\(a,q)=1}}^{q} \chi_1(a)\bar{\chi}_2(a) = \begin{cases}\phi(q) & \text{if } \chi_1 = \chi_2,\\ 0 & \text{if } \chi_1 \neq \chi_2.\end{cases}$$

Thus from (3),

$$\sum_{\substack{a=1\\(a,q)=1}}^{q} \left(\psi(x;q,a) - \frac{x}{\phi(q)}\right)^2 = \frac{1}{\phi(q)}\sum_{\chi}|\psi'(x,\chi)|^2.$$

As in the previous section, if χ is induced by χ_1, then

$$\psi'(x,\chi) = \psi'(x,\chi_1) + O((\log qx)^2).$$

Hence

$$\sum_{\substack{a=1\\(a,q)=1}}^{q} \left(\psi(x;q,a) - \frac{x}{\phi(q)}\right)^2 \ll (\log qx)^2 + \frac{1}{\phi(q)}\sum_{\chi}|\psi'(x,\chi_1)|^2.$$

Here the first term on the right is negligible, so that to prove (1) it suffices to show that

$$\sum_{q\le Q}\frac{1}{\phi(q)}\sum_{\chi}|\psi'(x,\chi_1)|^2 \ll xQ\log x.$$

If χ is primitive (mod q), then χ induces characters to moduli which are multiples of q; hence the left-hand side above is

$$\sum_{q\le Q}\sum_{\chi}{}^{*}|\psi'(x,\chi)|^2\sum_{k\le Q/q}\frac{1}{\phi(kq)}.$$

As in the previous section, the innermost sum is $\ll \phi(q)^{-1}\log(2Q/q)$. Hence it suffices to show that

(3) $$\sum_{q\le Q}\frac{1}{\phi(q)}\left(\log\frac{2Q}{q}\right)\sum_{\chi}{}^{*}|\psi'(x,\chi)|^2 \ll xQ\log x$$

for $x(\log x)^{-A} \le Q \le x$. We consider large and small q separately. From (2) we see that

$$\sum_{U<q\le 2U} \frac{1}{\phi(q)} \left(\log \frac{2Q}{q}\right) \sum_{\chi}^{*} |\psi(x, \chi)|^2 \ll (x^2 U^{-1} + Ux)(\log x)\left(\log \frac{2Q}{U}\right)$$

for $1 \le U \le Q$. Summing over $U = Q2^{-k}$, we find that

$$\sum_{Q_1<q\le Q} \frac{1}{\phi(q)} \left(\log \frac{2Q}{q}\right) \sum_{\chi}^{*} |\psi(x, \chi)|^2 \ll x^2 Q_1^{-1}(\log x)^2 + Qx \log x.$$

This suffices in (3), if $x(\log x)^{-A} \le Q \le x$ and $Q_1 = (\log x)^{A+1}$. By estimate (3) of §22,

$$\psi'(x, \chi) \ll x \exp(-c\sqrt{\log x})$$

for $q \le (\log x)^{A+1}$; hence the contribution of $q \le Q_1$ in (3) is

$$\ll Q_1(\log Q)x^2 \exp(-c\sqrt{\log x}) \ll x^2(\log x)^{-A} \ll Qx \log x.$$

Thus we have established (3), and the proof is complete.

30

REFERENCES TO OTHER WORK

The principal omission in these lectures has been the lack of any account of work on irregularities of distributions, both of the primes as a whole and of the primes in the various progressions to the same modulus q.

As regards irregularities in the distribution of the primes as a whole, the first point to be noted is that in this connection it is no longer possible to make inferences from the behavior of $\psi(x)$ to that of $\pi(x)$. It was proved by E. Schmidt in 1903, by relatively elementary arguments, that

$$\psi(x) - x = \Omega_{\pm}(x^{\frac{1}{2}}),$$

where the notation means that there exist arbitrarily large values of x for which

$$\psi(x) - x > cx^{\frac{1}{2}},$$

where c is some positive constant, and other arbitrarily large values of x for which

$$\psi(x) - x < -cx^{\frac{1}{2}}.$$

But the analogous problem for $\pi(x) - \operatorname{li} x$ was much more difficult. It had been conjectured, on numerical evidence, that $\pi(x) < \operatorname{li} x$ for all large x. This was disproved by Littlewood in 1914; he showed, in fact, that

$$\pi(x) - \operatorname{li} x = \Omega_{\pm}\left(\frac{x^{\frac{1}{2}} \log\log\log x}{\log x}\right).$$

Littlewood's proof[1] was divided into two cases, according as the Riemann hypothesis is true or false, the former being the difficult case. Owing to its indirect character, the proof did not make it

[1] See Ingham, Chap. 5, or Prachar, Chap. 7, §8.

possible to name a particular number x_0 such that $\pi(x) > \operatorname{li} x$ for some $x < x_0$. It was not until 1955 that such a number was found, namely by Skewes[2]; his number was $10_4(3)$, where $10_1(x) = 10^x$, $10_2(x) = 10^{10_1(x)}$, and so on.

Questions concerning the irregularity of distribution of the primes, as between one residue class to the modulus q and another, have been deeply studied in recent papers[3] on comparative prime number theory, by Turán and Knapowski. It is impossible to give any useful account of their work here, but one particular result may be mentioned as a sample. Suppose that, for each character $\chi \pmod q$, the function $L(s, \chi)$ has no zero in the rectangle

$$0 < \sigma < 1, \qquad |t| < \delta.$$

Then, if $a_1 \not\equiv a_2 \pmod q$, the difference

$$\psi(x; q, a_1) - \psi(x; q, a_2)$$

changes sign at least once in every interval

$$\omega \leq x \leq \exp(2\sqrt{\omega}),$$

provided ω is greater than a certain explicit function of q and δ. Some of their results are independent of any such unproved hypothesis. The work of Turán and Knapowski is based in part on some of the methods developed by Turán in his book *Eine neue Methode in der Analysis und deren Anwendungen* (Budapest, 1953).

The problem of finding an upper bound for the least prime in a given arithmetic progression has received a remarkably satisfactory solution (considering its inherent difficulty) at the hands of Linnik. He proved[4] that there exists an absolute constant C such that, if $(a, q) = 1$, there is always a prime $p \equiv a \pmod q$ satisfying $p < q^C$. The proof is difficult.

A subject that has attracted attention, but concerning which the known results leave much to be desired, is that of the behavior of $p_{n+1} - p_n$, where p_n denotes the nth prime. As regards a universal upper bound for this difference, the first result was found by Hoheisel, who proved that there exists a constant α, less than 1, such that $p_{n+1} - p_n = O(p_n^\alpha)$. The best result so far known is due to Ingham,[5] who showed that this estimate holds for any α greater than 38/61.

[2] *Proc. London Math. Soc.*, (3)**5**, 48–69 (1955).

[3] The main series consists of eight papers in *Acta Math. Hungaricae*, **13**(1962) and **14**(1963), and a sequel of three papers in *Acta Arithmetica* **9, 10, 11** (1964–1965), together with a paper in *J. Analyse Math.*, **14**(1965).

[4] See Prachar, Chap. 10.

[5] *Quarterly J. of Math.*, **8**, 255–266 (1937).

In both cases, what is actually proved is that

$$\pi(x + x^{\alpha}) - \pi(x) \sim \frac{x^{\alpha}}{\log x} \qquad \text{as } x \to \infty.$$

In a crude sense one can say, in view of the prime number theorem, that the average of $p_{n+1} - p_n$ is $\log p_n$. Erdös was the first to prove that there are infinitely many n for which $p_{n+1} - p_n$ is appreciably greater than $\log p_n$, and Rankin[6] proved that there are infinitely many n for which

$$p_{n+1} - p_n > c(\log p_n)\frac{(\log_2 p_n)(\log_4 p_n)}{(\log_3 p_n)^2},$$

where $\log_2 x = \log \log x$ and so on, and c is a positive constant. In the opposite direction, Bombieri and I proved recently[7] that there are infinitely many n for which

$$p_{n+1} - p_n < (0.46\ldots)\log p_n.$$

Of course, if the "prime twins" conjecture is true, there are infinitely many n for which $p_{n+1} - p_n = 2$.

There is a somewhat paradoxical situation in connection with the limit points of the sequence

$$\frac{p_{n+1} - p_n}{\log p_n}.$$

Erdös and Ricci (independently) have shown that the set of limit points has positive Lebesgue measure, and yet no number is known for which it can be asserted that it belongs to the set.

For references to other work in multiplicative number theory, one should consult, in the first place, the articles of Bohr and Cramér, and of Hua.

[6] *J. London Math. Soc.*, **13**, 242–247 (1938).

[7] *Proc. Royal Soc.* (*London*), **A, 293**, 1–18 (1966).

INDEX

S

T

V

W

Z